

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 3744.31 Seconds
(without alignments)
9989.844 Million cell updates/sec

Title: US-10-082-830-98

Perfect score: 863

Sequence: 1 atgggttggtcattcaact.....gtgagaccatgatagctg 863

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pt.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_on.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	613	71.0	162100	9	AC093297	AC093297 Homo sapi
C 2	613	71.0	176269	2	AC012663	AC012663 Homo sapi
C 3	72.2	8.4	154837	2	AL845425	AL845425 Danio rer
C 4	72.2	8.4	161204	2	AL845324	AL845324 Danio rer
C 5	71	8.2	183662	2	AC113205	AC113205 Mus muscu
C 6	69.8	8.1	151406	5	AL935208	AL935208 Zebrafish
C 7	67.6	7.8	6062	6	AX251789	AX251789 Sequence
C 8	67.6	7.8	66993	2	AC138074	AC138074 Homo sapi
C 9	67	7.8	146871	2	EX842689	EX842689 Danio rer
C 10	66.2	7.7	6195	6	AX345493	AX345493 Sequence
C 11	65.6	7.6	239339	2	EX548071	EX548071 Danio rer
C 12	65.4	7.6	1407	8	AJ592026	AJ592026 Arabidops
C 13	65.4	7.6	1434	8	AJ592058	AJ592058 Arabidops
C 14	65.4	7.6	3683	6	AX598999	AX598999 Sequence
C 15	65.2	7.6	1453	8	AJ591978	AJ591978 Arabidops
C 16	64	7.4	54345	3	AC084152	AC084152 Caenorhab
C 17	63.6	7.4	151900	9	AC107419	AC107419 Homo sapi
C 18	63.6	7.4	349980	6	AX344553	AX344553 Sequence
C 19	63.6	7.4	349980	6	AX344554	AX344554 Sequence
C 20	63.4	7.3	61864	3	CEY5088A	AL117200 Caenorhab
C 21	63	7.3	1228	8	AJ592059	AJ592059 Arabidops
C 22	63	7.3	172816	9	AC093899	AC093899 Homo sapi
C 23	62.8	7.3	7128	6	AX346461	AX346461 Sequence
C 24	62.8	7.3	40324	6	AX458634	AX458634 Sequence
C 25	62.6	7.3	104992	2	AC005504	AC005504 Plasmodiu
C 26	62.6	7.3	169546	2	AC004157	AC004157 Plasmodiu
C 27	62.6	7.3	250421	3	AE014849	AE014849 Plasmodiu
C 28	62.4	7.2	93791	2	AC138073	AC138073 Homo sapi
C 29	62.4	7.2	161699	9	AC107300	AC107300 Homo sapi
C 30	62.2	7.2	11147	6	AX598954	AX598954 Sequence
C 31	62.2	7.2	14147	6	AX767512	AX767512 Sequence
C 32	62.2	7.2	14147	6	AX795870	AX795870 Sequence
C 33	62.2	7.2	14147	6	AX822390	AX822390 Sequence
C 34	62.2	7.2	14147	6	AX826030	AX826030 Sequence
C 35	62.2	7.2	136335	8	AC005275	AC005275 Arabidops
C 36	62.2	7.2	195429	8	ATCHRIV8	AL161496 Arabidops
C 37	61.8	7.2	77242	5	AL591462	AL591462 Zebrafish
C 38	61.8	7.2	131682	9	AL672277	AL672277 Human DNA
C 39	61.8	7.2	13506	9	AC009950	AC009950 Homo sapi
C 40	61.6	7.1	12029	3	AE001372	AE001372 Plasmodiu
C 41	61.4	7.1	1184	8	AJ591982	AJ591982 Arabidops
C 42	61.4	7.1	157661	2	EX470134	EX470134 Danio rer
C 43	61.4	7.1	192182	5	EX649602	EX649602 Zebrafish
C 44	61.4	7.1	254050	3	PFA929358	AL929358 Plasmodiu
C 45	60.8	7.0	174273	9	CNS01RHZ	AL162759 Human chr

ALIGNMENTS

RESULT 1

AC093297/c 162100 bp DNA linear PRI 27-MAR-2002

LOCUS AC093297 Homo sapiens chromosome 5 clone RP11-53019, complete sequence.

DEFINITION AC093297

ACCESSION AC093297.3 GI:19747150

VERSION AC093297.3

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 162100)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

Pred. No. is the number of results predicted by chance to have a


```
AL845324
LOCUS       AL845324                161204 bp    DNA    linear    HTG 24-SEP-2003
DEFINITION   Danio rerio clone CH211-122L14, WORKING DRAFT SEQUENCE, 2 unordered
              pieces.
ACCESSION    AL845324
VERSION      AL845324.15   GI:35209193
KEYWORDS     HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       Danio rerio (zebrafish)
ORGANISM     Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE    1 (bases 1 to 161204)
AUTHORS      Sehra,H.
JOURNAL      Direct Submission
TITLE        Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-hel@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On Sep 24, 2003 this sequence version replaced gi:31616709.
COMMENT      ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-hel@sanger.ac.uk
              ----- Project Information
              Center project name: zc122l14
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 161092 bases at least Q40
              Consensus quality: 161096 bases at least Q30
              Consensus quality: 161100 bases at least Q20
              Insert size: 161104; sum-of-contigs
              Quality coverage: 166733; 4.8% error; agarose-fp
              coverage: 14.32x in Q20 bases; agarose-fp
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 2 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence.
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 140163: contig of 140163 bp in length
              * 140164 140263: gap of 100 bp
              * 140264 161204: contig of 20941 bp in length.
              Location/Qualifiers
              1..161204
              /organism="Danio rerio"
              /mol_type="genomic DNA"
              /db_xref="taxon:7955"
              /clone="CH211-122L14"
              /clone_lib="CHORI-211"
              1..140163
              /note="assembly fragment:01479"
              fragment_chain:1
              140264..161204
              /note="assembly fragment:04647"
              fragment_chain:1

misc_feature
1..140163
note="assembly fragment:01479"
fragment_chain:1

misc_feature
140264..161204
note="assembly fragment:04647"
fragment_chain:1

ORIGIN
Query Match      8.4%; Score 72.2; DB 2; Length 161204;
Best Local Similarity 32.3%; Pred. No. 0.0013;
Matches 260; Conservative 0; Mismatches 546; Indels 0; Gaps 0;

Qy 20 TTTAAGTTTATGAGATATATCCAGTGTGAATTTTGTAGCTGGTGTGATGATTTTACA 79
Db 145649 TTTTATTTTATTTCTATTTTCAATGTTCTCTTTAAATTAATATATATTTATAAA 145590
Qy 80 TTATATAGTATTACATTCATGATAGTCTCTCAGTAGATATCTCTCTATTTGTTTAAAT 139
Db 145589 TTATATATTTATTTATTTATTAACCTTTTATTTATTTATTTATTTATTTGCTTTTACT 145530
Qy 140 TTGGTTTGTCTTCATTTTCACTTATTTTAAACAGGCTCTTTAAATATACATCTTCTCG 199
Db 145529 TTTTGTATTTATTCATCTTTACTTTTACTTTTATTTATTTATTTATTTATTTATTTTACT 145470
Qy 200 TAGTGATGCTAGAAATGGAGTGGCTGGAATAAAGTGCGTGAATCTCTCAACTCTAG 259
Db 145469 TTTATGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 145410
Qy 260 TAAGATGCTAAACGTGTTTCTAAAGTGTTTATTTAACTATATATATGTCATTTTAA 319
Db 145409 TACTCTTTTACTTTATTTATTTTAAATTTATTTATTTTCTTTCTTTCTTTTATTTAT 145350
Qy 320 CAGCTCTTCACTTACTAGCAATTTATATCAGCAACACTGTGTATGTCAGACTTTTAA 379
Db 145349 TATTTAATTTATTTATTTTACTTTTCTTTTACTTTTACTTTTATTTATTTATTTATTTAT 145290
Qy 380 GTTTTCTTCACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 439
Db 145289 TAAATTTTATTTTATTTATTTATTTATTTATTTTCTTATTTATTTATTTATTTATTTA 145230
Qy 440 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 499
Db 145229 CTCTTACTTTTACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTACT 145170
Qy 500 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 559
Db 145169 TTTTACTTTTATTTAATATCAATTTATTTTATTTTATTTTATTTTACGCCACTTTTCTT 145110
Qy 560 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 619
Db 145109 TTTCTTTTAGATTTTCAGATATTTTCTTCTATTTTCTATTTTCTATTTTATTTATTTG 145050
Qy 620 NNNNNNNNGTCATCATTTATTTGTTGCTCATTTAAACAACTGTTAAATTTTCA 679
Db 145049 TATTTCAATTTATTTTATTTATTTGTTTACTTTTATTTTATTTAGATTTTATTTAAAT 144990
Qy 680 AATATCGACATTTATGATATTTATCTTCCAAAATTTAATAATTTGTCITTTTTCACAT 739
Db 144989 ATTTATTTTACTTAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 144930
Qy 740 TTAGCTTTTACTCAGCTGGAATTCATTTCTGTGTGCTGTGAGATAAATCTTTTTCAT 799
Db 144929 ATTTTATTTTACTTTTACTTTTACTTTTATTTTATTTTATTTTATTTTATTTTATTTAT 144870
Qy 800 GTTTTTTCCCATGAATAAATTTT 825
Db 144869 TTTATTTTATTTTATTTTATTT 144844

RESULT 4
```



```

* provided by the submittor
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 1269: contig of 1269 bp in length
* 1 1270 1369: gap of 100 bp
* 1 1370 3799: contig of 2430 bp in length
* 1 3800 3899: gap of 100 bp
* 1 3900 16887: contig of 12988 bp in length
* 1 16888 16987: gap of 100 bp
* 1 16988 44512: contig of 27525 bp in length
* 1 44513 44612: gap of 100 bp
* 1 44613 110938: contig of 66226 bp in length
* 1 110939 110938: gap of 100 bp
* 1 110939 156050: contig of 45112 bp in length
* 1 156051 156151 156350: gap of 100 bp
* 1 156151 183662: contig of 27512 bp in length.
FEATURES
      Location/Qualifiers
      1..183662
         /organism="Mus musculus"
         /mol_type="genomic DNA"
         /db_xref="taxon:10090"
         /clone="RP23-319G18"
         /clone_lib="RPC1-23 Female Mouse BAC"
      misc_feature
      1..1269
         /note="assembly_fragment"
         clone_end:SP6
         vector_side:left
      misc_feature
      1370..3799
         /note="assembly_fragment"
      misc_feature
      3900..16887
         /note="assembly_fragment"
      misc_feature
      16988..44512
         /note="assembly_fragment"
      misc_feature
      44613..110938
         /note="assembly_fragment"
      misc_feature
      110939..156050
         /note="assembly_fragment"
      misc_feature
      156151..183662
         /note="assembly_fragment"
         clone_end:T7
         vector_side:right
ORIGIN
Query Match      8.2%; Score 71; DB 2; Length 183662;
Best Local Similarity 35.7%; Pred. No. 0.002;
Matches 196; Conservative 0; Mismatches 334; Indels 19; Gaps 4;
QY      313 TTTTGACAGCTTTTCACCTTACTAGCAATTTATTATCAGCAACACTTGTTATTGTGAGA 372
      |||||
Db       57070 TTTCAAATAGGTTTGTTGTTAACTAGCAATTAGTCAGTCACACACTTGATCTCTAAGC 57011
      |||||
QY      373 CTTTAAAGTTTCATTCACTCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 432
      |||||
Db       57010 CTTTTAGGTTTTCATCTTCGTATAAATAGCTATCTCTTCGTGGTCTCATGGT 56951
      |||||
QY      433 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 492
      |||||
Db       56950 ATAATTCCCTTCCTGCTAATGAATTTAGAAAAGTTTCTATTCAATAGTCATTTCTCTTC 56891
      |||||
QY      493 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 552
      |||||
Db       56890 TGTGTGAAATTCATGTTAAGATTTTGTTCATCTTTGCACGTGATTTTCAACACAGATTT 56831
      |||||
QY      553 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 612
      |||||
Db       56830 ACTAGTTTGTGAATTTACCTTTCACCTACTCTATTATTATCAATTCATTTGGGCAGAACATCT 56771
      |||||
QY      613 NNNNNNNNNNNNGTGATCATATTTATGTTTGTCTATTTTAAACAACTGGTTAAAT 672
      |||||
Db       56770 TCAGTT-----TGGTCATCATATGCTGAICTCAITTTAAAAAATTCAGCATAGG 56721
      |||||
QY      673 ATTTTCAC--AAATCGACATTATGATATATTATCTTCCAAAATTTTAAATAATTTGTGCTTT 730
      |||||

```

Db	56720	AGTACTTTTAACTCATGTTACAGATATTATCTTCCAAATTTTGTAAATTTT	---	CCT	56664
Qy	731	TTTCACATTTTAGTCTTTAGCTTCAGCTGGAATTCATTTCTGTGCTGGTGTGAGATAAG-	789		
Db	56663	CTTTGCATTTGATTTTGTCTGCATCTCGAGTTGATTTTGGTGCCTGGTGTATGAAGG	56604		
Qy	790	---TCCTTTTCATGTTTTTCCCTATGAATAAATTAATTCCTTCCTGTTATTCGAGGTAGCT	846		
Db	56603	GTCGATTTTCTGTGTTTTCCCGTGACTAATAATGAGCTCATCTTATCTATGTAGACAGCT	56544		
Qy	847	GAGACCAAT	855		
Db	56543	GAGACCGAT	56535		
RESULT 6	AL935208	151406 bp	DNA	linear	VRT 17-JUL-2003
LOCUS	Zebrafish	DNA sequence from clone CH211-241N17,	complete	sequence.	
DEFINITION	AL935208				
ACCESSION	AL935208.8	GI:32949784			
VERSION	HTG.				
KEYWORDS	Danio rerio (zebrafish)				
SOURCE	Danio rerio				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.				
REFERENCE	1 (bases 1 to 151406)				
AUTHORS	Phillimore, B.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUL-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 17, 2003 this sequence version replaced gi:32398511.				
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: zfish-help@sanger.ac.uk				

During sequence assembly a bias is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C elegans/wormpep> Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr.' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dr.' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see <http://www.sanger.ac.uk/Projects/D rerio/fishmask.shtml> CH211-241N17 is from a CHORI-211 BAC library VECTOR: pTARBAC2.1

Location/Qualifiers
1..151406

FEATURES
source

```
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-241N17"
/clone_lib="CHORI-211"

ORIGIN
Query Match      8.1%; Score 69.8; DB 5; Length 151406;
Best Local Similarity 32.6%; Pred. No. 0.0033;
Matches 250; Conservative 0; Mismatches 515; Indels 1; Gaps 1;

QY 1 ATGGGTTGGTCATCAACTTAAGTTATGAGATATATCCATGTCGTAATTTTGTAGCTG 60
Db 67975 ATGGTTGGTTAAAGTTTTTTTATGTTTATTAATAAAATATTTCTTTAAATGTTTTT 68034

QY 61 TGGTTTGATGATTTTACATATATAGTATTACATTCATCCATGGATGTTCTCAGTAGATA 120
Db 68035 TTATATTTTATTAATCAATTTATTAATATATTTATTTATTTATTTTGTGTTT 68094

QY 121 TCCCTCATGTTTAAACATTTGCGTCTCTCATTTGACCTATTTTAAACAGGCTCT 180
Db 68095 TATGCAATTTCTTTAAATTAATTTAAATATTTATTTATTTTATTTTAAATAT 68154

QY 181 TTAATATACACTTCTCTGATGCTATGCTAGAAATCGAGTGGCTGGAATAAAGGCT 240
Db 68155 TTTATTTCTTTAAATTCATTTATTTAAATTTAAATTTATTTTGTGTTATTTAT 68214

QY 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACTGTTTTCTAAA-GTGTTTTATTTAAC 299
Db 68215 TTATATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 68274

QY 300 TATTATATGCAATTTTGAACAGCTCTTTCACCTACTAGCAATTTATATCAGCAACT 359
Db 68275 TTTTAAATCTCTTTGTTTAACTATTTATTTTAAATTTATTTTATTTAAATATTTT 68334

QY 360 TGTATTGTGACAGCTTTTAAAGTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 419
Db 68335 TATTTTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 68394

QY 420 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 479
Db 68395 AATTATTTATTTTGTATTTATTTTAAATTTCTTTTATTTGTAATTTTATGCGTTT 68454

QY 480 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 539
Db 68455 TTTTCTCTTAAATGTTTATTTAAATTTATTTATTTATTTATTTATTTATTTAAAT 68514

QY 540 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 599
Db 68515 TTTATTTTCATAATTTTCATTTATTTTATTTAAATTTTATTTGTAATTTTTCATTTT 68574

QY 600 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 659
Db 68575 TTTATTTATTTCAATTTAAATTTAAATTTTAAATTTTATTTATTTTATTTATTTATTT 68634

QY 660 AACCTGTTAAATTTTCAAAATCGACATTTATGATATATTTCTTCCAAAATTTTAAATA 719
Db 68635 ATTTTAGTATTTTTCATTTAAATTTAAATTTAAATTTTATTTATTTTATTTATTTATTT 68694

QY 720 ATTTTGTCTTTTTCACATTTTAGTCTTTAGCTAGCTGGAATCA 765
Db 68695 ATTTTATATAATTTTATTTTACTTTTATTTATTTATTTATTTTCTTATATTTCA 68740

RESULT 7
AX251789
LOCUS AX251789 5062 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 50 from Patent WO0168911.
ACCESSION AX251789
VERSION AX251789.1 GI:15985144
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
```

artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the cell cycle
JOURNAL Patent: WO 0168911-A 50 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
1. 6062
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 7.8%; Score 67.6; DB 6; Length 6062;
Best Local Similarity 32.8%; Pred. No. 0.018;
Matches 270; Conservative 0; Mismatches 552; Indels 1; Gaps 1;

QY 20 TTTAAGTTTATGAGATATATCCATGTTTGAATTTTGTAGCTGTTGATGATTTTACA 79
Db 5122 TTTAGTCTGCTATATATATAGTAATTTTAAATGTAAGTTTGTAGTTTATTTT 5181

QY 80 TTATATAGTATTACATTCATGATAGTTCTCAGTAGATAATCTCTATTTGTTTAA 139
Db 5182 TTGATTATAATTTATATGTTTATTTGTTTAAATATTTGTTTGTAGTTTAA 5241

QY 140 TTGCTTCTCTCTCATTTGACCTATTTTAAACAGGCTCTTTAAATATACACTTCT 199
Db 5242 ATTTTAAAAATTTTAAATTTTATTTTATTTATTTTATTTTATTTTAAATTT 5301

QY 200 TAGTCTATGCTAGAAATGAGTGGCTGGAATAAAGTGGCTGATCATCTCAACTCT 259
Db 5302 AAGTGGGTATTTAAT-TGCTTTTGTAAATAGAAATAGGTAATTTATTTAGTTC 5360

QY 260 TAAGATGTCAAACTGTTTCTTAAAGTGTTTTATTTTAACTATTTATATGCAATTT 319
Db 5361 TTTTGTGTTAGTTATTTAGAAAGTTTATTTAGTTTATTTTATTTTAAATTTT 5420

QY 320 CAGCTCTTCTACTACTAGCAATTTATATCAGCAACTGTTTATTTGTCAGACTTTAA 379
Db 5421 TTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTTTATTTTATTT 5480

QY 380 GTTTTCATCTACTGNN 439
Db 5481 GTAGTAGTTTGTGTCGGTAAATTTATTTTATTTTATATCTGATTTAGAGTAT 5540

QY 440 NNN 499
Db 5541 TTTTAAATTTTAAATTTAGTTAGTTTGTTTTAAAGTTTAAATTTTGTGAAGAA 5600

QY 500 NNN 559
Db 5601 ATTTGTATGATATTTTATTTTATTTTGTGTTAAATTTGTTATTTGTTTATTT 5660

QY 560 NNN 619
Db 5661 TTAGTAAGTTTGTGTTTATTTTATTTTATTTTATTTATTTATTTTATTTTATTT 5720

QY 620 NNNNNNGTATCATATTTATGTTTGTCTCATTTAAATAAACTGTTTAAATTTTCA 679
Db 5721 TTTTAAAGTTTGTAGTTTAAATTTATTTTAAAGTTTAAAGTTTAAAGTTTAA 5780

QY 680 AAATCGACATTTATGATATTTATCTTCCAAAATTTTAAATTTTGTCTTTTTCACAT 739
Db 5781 AATTAGGTAGGTATTTTATTTTATTTTATTTATTTATTTTATTTTATTTTATTT 5840

QY 740 TTAGTCTTTAGCTAGCTGGAATTTCTGTTGTTGTTGTTGTTGTTGTTGTTTCA 799
Db 5841 TAGTAGAGAGTTTGTGTAATTTTATTTTATTTTAAAGAGAGATAAATTTAAAG 5900

QY 800 GTTTTTCCTATGAATAAATTTTCTCTCTGTTTATTTGAGGT 842
Db 5901 TTTTAAAGTTAAGATTTTGTGTTTATTTTGTAGTT 5943

```

RESULT 8
AC138074
LOCUS
DEFINITION
    Homo sapiens chromosome 18 clone RP13-907D8 map 18, LOW-PASS
    SEQUENCE SAMPLING.
ACCESSION
    AC138074.1 GI:26553381
VERSION
    HTG; HTGS PHASE0.
KEYWORDS
    Homo sapiens (human)
SOURCE
    Homo sapiens
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
        1 (bases 1 to 66993)
REFERENCE
    Birren, B., Nusbaum, C. and Lander, E.
    Homo sapiens chromosome 18, clone RP13-907D8
    Unpublished
    2 (bases 1 to 66993)
AUTHORS
    Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
    Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
    Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
    Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
    Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
    Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
    Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
    Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
    Landblad-Ton, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
    Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,
    Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
    Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
    Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
    Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R.,
    Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
    Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Toham, K.,
    Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
    Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
    Direct Submission
    Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    All repeats were identified using RepeatMasker:
    Smith, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html
    ----- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
    ----- Project Information
    Center project name: L28874
    Center clone name: 907_D_8
    -----
    * NOTE: This record contains 55 individual
    * sequencing reads that have not been assembled into
    * contigs. Runs of N are used to separate the reads
    * and the order in which they appear is completely
    * arbitrary. Low-pass sequence sampling is useful for
    * identifying clones that may be gene-rich and allows
    * overlap relationships among clones to be deduced.
    * However, it should not be assumed that this clone
    * will be sequenced to completion. In the event that
    * the record is updated, the accession number will
    * be preserved.
    *
    * 1 1116: contig of 1116 bp in length
    * 1117 1216: gap of 100 bp
    * 1217 2336: contig of 1120 bp in length
    * 2337 2436: gap of 100 bp
    * 2437 3561: contig of 1125 bp in length
    * 3562 3661: gap of 100 bp
    * 3662 4781: contig of 1120 bp in length
    * 4782 4881: gap of 100 bp
    * 4882 6001: contig of 1120 bp in length
    * 6002 6101: gap of 100 bp
    *
    * 6102 7244: contig of 1143 bp in length
    * 7245 7345: gap of 100 bp
    * 7346 8458: contig of 1114 bp in length
    * 8459 8559: gap of 100 bp
    * 8560 9640: contig of 1082 bp in length
    * 9641 9740: gap of 100 bp
    * 9741 10846: contig of 1106 bp in length
    * 10847 10946: gap of 100 bp
    * 10947 12103: contig of 1157 bp in length
    * 12104 12203: gap of 100 bp
    * 12204 13200: contig of 1117 bp in length
    * 13201 13321: gap of 100 bp
    * 13322 14520: contig of 1100 bp in length
    * 14521 14620: gap of 100 bp
    * 14621 15773: contig of 1153 bp in length
    * 15774 15873: gap of 100 bp
    * 15874 16991: contig of 1118 bp in length
    * 16992 17091: gap of 100 bp
    * 17092 18240: contig of 1149 bp in length
    * 18241 18340: gap of 100 bp
    * 18341 19476: contig of 1136 bp in length
    * 19477 19576: gap of 100 bp
    * 19577 20721: contig of 1145 bp in length
    * 20722 20821: gap of 100 bp
    * 20822 21942: contig of 1121 bp in length
    * 21943 22042: gap of 100 bp
    * 22043 23141: contig of 1099 bp in length
    * 23142 23241: gap of 100 bp
    * 23242 24372: contig of 1131 bp in length
    * 24373 25627: contig of 1155 bp in length
    * 25628 25727: gap of 100 bp
    * 25728 26888: contig of 1161 bp in length
    * 26889 28030: contig of 1042 bp in length
    * 28031 28131: gap of 100 bp
    * 28132 29303: contig of 1173 bp in length
    * 29304 29403: gap of 100 bp
    * 29404 30554: contig of 1151 bp in length
    * 30555 30654: gap of 100 bp
    * 30655 31751: contig of 1097 bp in length
    * 31752 31851: gap of 100 bp
    * 31852 32966: contig of 1115 bp in length
    * 32967 33066: gap of 100 bp
    * 33067 34113: contig of 1047 bp in length
    * 34114 34213: gap of 100 bp
    * 34214 35329: contig of 1116 bp in length
    * 35330 35429: gap of 100 bp
    * 35430 36574: contig of 1145 bp in length
    * 36575 36674: gap of 100 bp
    * 36675 37828: contig of 1154 bp in length
    * 37829 37928: gap of 100 bp
    * 37929 39063: contig of 1135 bp in length
    * 39064 39163: gap of 100 bp
    * 39164 40289: contig of 1126 bp in length
    * 40290 40389: gap of 100 bp
    * 40390 41471: contig of 1082 bp in length
    * 41472 41571: gap of 100 bp
    * 41572 42634: contig of 1063 bp in length
    * 42635 42734: gap of 100 bp
    * 42735 43904: contig of 1170 bp in length
    * 43905 44004: gap of 100 bp
    * 44005 45128: contig of 1124 bp in length
    * 45129 45228: gap of 100 bp
    * 45229 46358: contig of 1130 bp in length
    * 46359 46458: gap of 100 bp
    * 46459 47595: contig of 1137 bp in length
    * 47596 47695: gap of 100 bp
    * 47696 48785: contig of 1090 bp in length
    * 48786 48885: gap of 100 bp
    * 48886 49960: contig of 1075 bp in length
    * 49961 50060: gap of 100 bp
    * 50061 51132: contig of 1072 bp in length

```

```

* 51133 51232: gap of 100 bp
* 51233 52356: contig of 1124 bp in length
* 52357 52456: gap of 100 bp
* 52457 53556: contig of 1100 bp in length
* 53557 53656: gap of 100 bp
* 53657 54782: contig of 1126 bp in length
* 54783 54882: gap of 100 bp
* 54883 56008: contig of 1126 bp in length
* 56009 56108: gap of 100 bp
* 56109 57307: contig of 1199 bp in length
* 57308 57407: gap of 100 bp
* 57408 58529: contig of 1122 bp in length
* 58530 58629: gap of 100 bp
* 58630 59799: contig of 1170 bp in length
* 59800 59899: gap of 100 bp
* 59900 60983: contig of 1084 bp in length
* 60984 61083: gap of 100 bp
* 61084 62193: contig of 1110 bp in length
* 62194 62293: gap of 100 bp
* 62294 63397: contig of 1104 bp in length
* 63398 63497: gap of 100 bp
* 63498 64574: contig of 1077 bp in length
* 64575 64674: gap of 100 bp
* 64675 65800: contig of 1126 bp in length
* 65801 65900: gap of 100 bp
* 65901 65993: contig of 1093 bp in length.

FEATURES
    source
    1. .66993
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="18"
    /map="18"
    /clone="RP13-907D8"
    /clone_lib="RPC1-13 Human Female BAC"

ORIGIN
Query Match      7.8%; Score 67.6; DB 2; Length 66993;
Best Local Similarity 33.5%; Pred. No. 0.0098;
Matches 278; Conservative 0; Mismatches 553; Indels 0; Gaps 0;

Qy 6 TTTCGTCATCACTTAAAGTTATGAGATATATCCATGTTGAATTTTGTAGCTGGTT 65
Db 56412 TTTTNNNTTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56471

Qy 66 TGATGATTTTACATTTATAGTATATACATTCATCCATGATAGTCTCAGTAGATAATCCTC 125
Db 56472 TTTTNTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 56531

Qy 126 CTATGTTTAAACATTTGCGTGTCTCTCATTTTGACCTATTTTAAACAGCCCTCTTTAAA 185
Db 56532 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56591

Qy 186 TATACATCTCTGCTAGTCTGCTAGATAATGAGTGGCTGGAATAAAGTGGCTGAATC 245
Db 56592 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56651

Qy 246 ATCTTCAACTCTAGTAGATGTCACAACTGTTTCTAAAGTGTTTTATTTTAACTATTAT 305
Db 56652 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56711

Qy 306 ATGTCAAATTTGACAGCTCTTCACTTACTAGCAATTTATATACAGCACTTGTAT 365
Db 56712 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56771

Qy 366 TGTTCAGACTTTTAAAGTTTTCATTCAGCTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 425
Db 56772 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56831

Qy 426 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 485
Db 56832 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56891

```

```

Qy 486 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 545
Db 56892 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 56951

Qy 546 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 605
Db 56952 NNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 57011

Qy 606 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 665
Db 57012 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57071

Qy 666 GTTAAATATTTCAAAATCGACATATGATATATATCTTCCAAAATTTTAAATATTTTG 725
Db 57072 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57131

Qy 726 TCTTTTTCACATTTTACGCTTTAGCTCAGCTGGAATTCATTTCTGTGTGTGTGAGA 785
Db 57132 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57191

Qy 786 TAAGTCCTTTTTCATGTTTTTCCCTATGAAATAAATAATTTCCCTCTCTATT 836
Db 57192 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57242

RESULT 9
BX842689/c 146871 bp DNA linear HTG 03-DEC-2003
LOCUS Danio rerio clone DKEY-96D22, WORKING DRAFT SEQUENCE, 11 unordered
DEFINITION pieces.
ACCESSION BX842689
VERSION BX842689.1 GI:38677860
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 146871)
Burton, J.
Direct Submission
Submitted (02-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zk96D22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 143882 bases at least Q40
Consensus quality: 144621 bases at least Q30
Consensus quality: 145107 bases at least Q20
Insert size: 145871; sum-of-contigs
Insert size: 162937; 4.9% error; agarose-fp
Quality coverage: 5.37X in Q20 bases; sum-of-contigs Quality
coverage: 4.81X in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 19996: contig of 19996 bp in length
* 19997 20096: gap of 100 bp
* 20097 44027: contig of 23931 bp in length

```

[illegible]

[illegible]

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 2958.17 Seconds
(without alignments)
8711.839 Million cell updates/sec

Title: US-10-082-830-98

Perfect score: 863

Sequence: 1 atgggttttggtcattcaact.....gtgagaccaatgatagctg 863

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_estc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_esthum.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.2	10.6	986	12	BG116503 602317590
2	89.8	10.4	679	13	BX488216 DKFZp686P
3	78	9.0	1200	13	EX437758 BX437758
4	78	9.0	1392	29	CG757503 P052-4-C0

5	72.6	8.4	1811	29	CG753732
6	71.2	8.3	1288	29	CG744915
7	69.8	8.1	1221	28	CC301561
8	69.2	8.0	1380	29	CG744815
9	68.2	8.0	1389	29	CG747695
10	68.6	7.9	1531	29	CG748014
11	68.4	7.9	1297	29	CG758143
12	68.2	7.9	1491	29	CG753221
13	67.2	7.8	1377	29	CG749971
14	66.8	7.7	1376	29	CG747831
15	66.6	7.7	1275	29	CG754010
16	66.6	7.7	1353	29	CG74812
17	66	7.6	1566	29	CG757757
18	65.6	7.6	1135	29	CNS033GQ
19	65.6	7.6	1434	28	CC187638
20	65.2	7.6	1361	29	CG744327
21	65	7.5	973	29	CNS0711E
22	65	7.5	1139	28	AQ897537
23	65	7.5	1313	29	CG751144
24	65	7.5	1352	28	CC288594
25	64.8	7.5	1225	29	CG745927
26	64.8	7.5	1331	29	CG749578
27	64.6	7.5	1260	29	CG751608
28	64.6	7.5	1283	29	CG749744
29	64.4	7.5	1373	29	CG750869
30	64.4	7.5	1433	29	CG745119
31	64.2	7.4	1101	29	CNS003BD
32	64.2	7.4	1298	29	CG756607
33	64.2	7.4	1362	29	CG757918
34	64.2	7.4	1528	29	CG753854
35	64	7.4	1324	29	CG746828
36	63.8	7.4	1254	29	CG748584
37	63.8	7.4	1626	14	CF238805
38	63.6	7.4	980	29	CNS0141E
39	63.4	7.3	1224	29	CG757393
40	63.4	7.3	1235	29	CG750878
41	63.4	7.3	1269	29	CG757211
42	63.4	7.3	1426	28	CC231597
43	63.2	7.3	1018	12	BM416001
44	63.2	7.3	1063	13	EX414736
45	63	7.3	1211	29	CG747324

ALIGNMENTS

RESULT 1
BG116503
LOCUS 602317590F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417870 5',
DEFINITION mRNA sequence.
ACCESSION BG116503.1 GI:12610009
VERSION BG116503
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 986)
AUTHORS NIH-MGC <http://mgc.ncbi.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10151 row: c column: 23
High quality sequence stop: 634.

CG753732 P048-4-G0
CG744915 P037-3-F0
CC301561 CH261-13K
CG744815 P037-3-B0
CG747695 P041-2-D0
CG748014 P041-4-B0
CG758143 P053-3-B1
CG753221 P048-2-A0
CG749971 P044-2-E0
CG747831 P041-3-B0
CG754010 P049-2-C0
CG74812 P037-3-B0
CG757757 P053-1-D0
AL228115 Tetradon
CC187638 CH261-98P
CG744327 P036-4-E0
AL425064 clone BAO
AQ897537 HS 3153 A
CG751144 P045-1-E0
CC288594 CH261-67F
CG745927 P039-1-A0
CG749578 P043-4-D1
CG751608 P046-1-C1
CG749744 P044-1-C1
CG750869 P045-2-E1
CG745119 P037-4-G0
AL064091 Drosophil
CG756607 P051-4-C0
CG757918 P053-2-A0
CG745927 P039-1-A0
CG753854 P049-1-D0
CG746828 P040-1-G1
CG748584 P042-3-B0
CF238805 AGENCOURT
AL103580 Drosophil
CG757393 P052-3-G0
CG750878 P045-2-F0
CG757211 P052-2-G0
CC231597 CH261-36A
BM416001 OP21084 M
EX414736 BX414736
CG747324 P040-4-D1

BG116503 986 bp mRNA linear EST 30-JAN-2001
602317590F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417870 5',
mRNA sequence.

ACCESSION BG116503.1 GI:12610009
VERSION BG116503
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 986)

NIH-MGC <http://mgc.ncbi.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10151 row: c column: 23
High quality sequence stop: 634.

[illegible]

Query Match 8.0%; Score 69.2; DB 29; Length 1389;
 Best Local Similarity 33.2%; Pred. No. 0.00087;
 Matches 276; Conservative 0; Mismatches 555; Indels 0; Gaps 0;

QY 6 TTGGTCATTCACATTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGGTT 65
 DB 519 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 578

QY 66 TGATGATTTTACATATATAGTATATACATCCAGGATAGTCTCAGTAGATAATCCCTC 125
 DB 579 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 638

QY 126 CTATGTTTAAACATTCGGTTGCTCTCATTTTGACCTATTTTAAACAGGCCCTTTAAA 185
 DB 639 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 698

QY 186 TATACACTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
 DB 699 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 758

QY 246 ATCTTCACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305
 DB 759 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 818

QY 306 ATGTCATTTTGAACAGCTCTTTCACCTACTAGCAATTTATATCAGCAACACTGTTAT 365
 DB 819 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 878

QY 366 TGCAGACTTTTAAAGTTTCACTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 425
 DB 879 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 938

QY 426 NNN 485
 DB 939 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 998

QY 486 NNN 545
 DB 999 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1058

QY 546 NNN 605
 DB 1059 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1118

QY 606 NNN 665
 DB 1119 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1178

QY 666 GTTAAATATTCACAAATCGACATATGATATATATCTCCAAATTTTAAATTTTG 725
 DB 1179 NNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1238

QY 726 TCTTTTTCACATTTTAGCTTTAGCTCAGCTGGAATTCATTTCTGTGTGGTGTGAGA 785
 DB 1239 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1298

QY 786 TAAGTCTTTTTCAGTTTTTCCCTAGAAATAAATATTTCTCTCTGATTT 836
 DB 1299 NTTTTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1349

RESULT 10
 CG748014/c
 LOCUS
 DEFINITION P041-4-B02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
 genomic survey sequence.
 ACCESSION CG748014
 VERSION CG748014.1 GI:37968940
 KEYWORDS GSS.
 SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (bases 1 to 1531)
 Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
 Buntjer,J., van der Meulen,M. and Sommer,R.J.
 An integrated physical and genetic map of the nematode Pristionchus
 pacificus
 Mol. Genet. Genomics 269 (5), 715-722 (2003)
 22835951
 12884007
 Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@uebingen.mpg.de
 Class: BAC ends.

Location/Qualifiers
 1. .1531
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strains="California"
 /db_xref="taxon:54126"
 /clone_lib="Ppa EcoRI BAC Library"
 /note="The library was generated by a partial digest of
 the genomic DNA with EcoRI and cloning into the BAC
 vector."

FEATURES

source

ORIGIN

Query Match 7.9%; Score 68.6; DB 29; Length 1531;
 Best Local Similarity 33.6%; Pred. No. 0.0011;
 Matches 279; Conservative 0; Mismatches 552; Indels 0; Gaps 0;

QY 6 TTGTGCTTCACTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGGTT 65
 DB 1320 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1261

QY 65 TGATGATTTTACATATATAGTATATACATCCAGGATAGTCTCAGTAGATAATCCCTC 125
 DB 1260 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1201

QY 126 CTATGTTTAAACATTTGGCTTCTCTCATTTTGACCTATTTTAAACAGGCCCTTTAAA 185
 DB 1200 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1141

QY 186 TATACACTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
 DB 1140 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1081

QY 246 ATCTTCACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305
 DB 1080 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1021

QY 306 ATGTCATTTTGAACAGCTCTTTCACCTACTAGCAATTTATATCAGCAACACTGTTAT 365
 DB 1020 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 961

QY 366 TGTACAGACTTTTAAAGTTTTCATCTGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 425
 DB 960 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 901

QY 426 NNN 485
 DB 900 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 841

QY 486 NNN 545
 DB 840 NTTTTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 781

QY 546 NNN 605
 DB 780 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 721

[illegible]

RESULT 14
CG747831
LOCUS
CG747831 1376 bp DNA linear GSS 24-OCT-2003
DEFINITION P041-3-B05.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.

```

FEATURES             .
  source
    Class: BAC ends.
    Location/Qualifiers
      1. .i376
        /organism="Pristionchus pacificus"
        /mol_type="genomic DNA"
        /strain="California"
        /db_xref="taxon:54126"
        /clone_lib="ppa EcoRI BAC Library"
        /note="The library was generated by a partial digest of
        the genomic DNA with EcoRI and cloning into the BAC
        vector."

ORIGIN
Query Match          7.7%;   Score 66.8;   DB 29;   Length 1376;
Best Local Similarity 32.9%;   Pred. No. 0.0025;
Matches 273;   Conservative 0;   Mismatches 557;   Indels 0;   Gaps 0;

7  TTGGTCATTCACACTTTAACTTTATCAGAGATATATCCATGTTGAATTTTGTAGCTGGTTT 66

```

[illegible]

RESULT 15	CG754010	1276 bp	DNA	linear	GSS 24-OCT-2003
LOCUS	CG754010				
DEFINITION	P043-2-C03.ya Ppa EcoRI BAC Library				<i>Pristionchus pacificus</i> genomic, genomic survey sequence.
ACCESSION	CG754010				
VERSION	CG754010.1	GI:37979071			
KEYWORDS	GSS.				
SOURCE	<i>Pristionchus pacificus</i>				
ORGANISM	<i>Pristionchus pacificus</i>				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;				
	Neodiplogasteridae; <i>Pristionchus</i> .				
REFERENCE	1 (bases 1 to 1276)				
AUTHORS	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.				
TITLE	An integrated physical and genetic map of the nematode <i>Pristionchus pacificus</i>				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 434.526 Seconds
(without alignments)
8437.229 Million cell updates/sec

Title: US-10-082-830-98
Perfect score: 863
Sequence: 1 atgggtttgtcattcaact.....gttgagaccatgatagctg 863

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	630	73.0	863	6	ABK93492 Human bre
2	392	45.4	392	6	ABK93491 Human bre
3	89.8	10.4	229	4	AAL26040 Human bre
C 3	89.8	10.4	233	4	AAL16838 Human bre
C 4	89.8	10.4	256	4	AAL17847 Human bre
C 5	89.8	10.4	258	4	AAL17770 Human bre
6	89.8	10.4	270	4	AAL10695 Human bre
7	89.8	10.4	298	4	AAL10772 Human bre
C 8	89.8	10.4	317	4	AAL17429 Human bre
9	89.8	10.4	332	4	AAL09537 Human bre
10	89.8	10.4	325	4	AAL21979 Human bre
C 11	85	9.8	525	4	AAL08295 Human bre
C 12	77.8	9.0	248	4	AAL13109 Human bre
C 13	71.6	8.3	534	4	AAL10199 Human gen
14	67.6	7.8	6062	6	AAS61094 Human imm
15	66.2	7.7	6195	6	ABL32591 Human imm
16	65.4	7.6	3683	7	ABZ10199 Haematopo
17	62.8	7.3	7128	6	ABL33559 Human imm
18	62.8	7.3	40324	6	ABQ67150 Human ang
19	62.2	7.2	11147	7	ABZ10154 Haematopo
20	62.2	7.2	14147	9	ABZ54226 Pretreat
21	62.2	7.2	14147	9	ABE84164 Human lym
22	60.6	7.0	9964	6	ABL32098 Human imm
23	60.4	7.0	11147	7	ABZ10008 Haematopo

24	60.4	7.0	14147	4	AAS46744 Tumour su
25	60.4	7.0	14147	6	ABK33956 Human DNA
26	60.4	7.0	14147	7	ADA20387 Prostate
27	60.4	7.0	14147	7	ADA84194 Human ren
28	60.4	7.0	14147	9	ADB54098 Pretreat
29	60.4	7.0	14147	9	ADE84088 Human lym
30	59.6	6.9	7442	4	AAS46686 Tumour su
31	59.6	6.9	15732	4	AAS45388 Chemical
32	59.6	6.9	15732	6	ABK28233 DNA trans
33	59.4	6.9	6171	6	ABL33010 Human imm
34	59.2	6.9	3683	7	ABZ10053 Haematopo
35	58.8	6.8	5327	4	AAS45356 Chemical
36	58.8	6.8	5327	6	ABK28189 DNA trans
37	58.8	6.8	9539	4	AAS45347 Chemical
38	58.8	6.8	9539	6	ABK28180 DNA trans
39	58.6	6.8	12237	6	ABL34358 Human imm
40	58.2	6.7	6132	6	ABL32863 Human imm
41	58	6.7	12356	4	AAS46510 Tumour su
42	57.6	6.7	6000	6	ABL34305 Human imm
43	57.6	6.7	6292	4	AAS46735 Tumour su
44	57.6	6.7	10467	6	ABL49302 Human pol
45	57.2	6.6	5362	6	ABK28413 DNA trans

ALIGNMENTS

RESULT 1
ABK93492
ID ABK93492 standard; CDNA; 863 BP.
XX
AC ABK93492;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human breast specific nucleic acid, BSNA #98.
XX
KW Human; ss; breast specific nucleic acid; BSNA; breast cancer;
KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
XX
OS Homo sapiens.
XX
PN WO200236807-A2.
XX
PD 10-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-US046888.
XX
PR 27-OCT-2000; 2000US-0243802P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
XX
DR WPI; 2002-463415/49.
XX
PT New breast-specific nucleic acids and polypeptides, useful for
PT identifying, diagnosing, monitoring, staging, imaging, and treating
PT breast cancer and non-cancerous disease states in breast tissues.
XX
PS Claim 1; Page 191; 281pp; English.
XX
CC The invention relates to breast specific nucleic acids (BSNA) and breast-
CC specific polypeptides (BSP). Also included are a method for determining
CC the BSNA in a sample, a vector comprising a BSNA, a host cell comprising
CC the vector, a method for producing a polypeptide encoded by a BSNA, an
CC anti-BSP antibody and a method for determining the presence of a BSP in a
CC sample. The breast-specific nucleic acids, polypeptides and compositions
CC comprising them are useful for identifying, diagnosing, monitoring and
CC staging, imaging, and treating breast cancer, mammary tumour and non-
CC cancerous disease states in breast tissue; for identifying breast tissue;
CC for monitoring, identifying and/or designing agonists and antagonists of
CC the polypeptides; in gene therapy; in producing transgenic animals and

CC cells; for producing engineered breast tissue for treatment and research;
 CC and as elements in an array or computer program for pattern recognition
 CC of breast disorders. The nucleic acids may be used as hybridisation
 CC probes to detect, characterise and quantify hybridising nucleic acids in,
 CC and isolate hybridising nucleic acids from, both genomic and transcript-
 CC derived nucleic acid samples. The BSP protein may be used in a vaccine
 CC composition for raising an immune response against breast cancer. The
 CC present sequence is BSNA cDNA of the invention
 XX
 SQ Sequence 863 BP; 173 A; 94 C; 98 G; 265 T; 0 U; 233 Other;

Query Match 73.0%; Score 630; DB 6; Length 863;
 Best Local Similarity 100.0%; Pred. No. 1.8e-121;
 Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGGTCATTCAACTTTAAGTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 DB 1 ATGGGTTGGTCATTCAACTTTAAGTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 QY 61 TGGTTGATGATTTTACATTTATAGTATTTACATTCATGATGATTTTGTAGCTG 120
 DB 61 TGGTTGATGATTTTACATTTATAGTATTTACATTCATGATGATTTTGTAGCTG 120
 QY 121 TCCCTCCTATTGTTTAAACATTTGCGTCTCTCATTTTGGACCTATTTTAAACAGGCTCT 180
 DB 121 TCCCTCCTATTGTTTAAACATTTGCGTCTCTCATTTTGGACCTATTTTAAACAGGCTCT 180
 QY 181 TTAATATACACTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 DB 181 TTAATATACACTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 241 GAATCATCTCAACTCTAGTAAAGTCAAACTGTTTCTAAAGTGTATTTTAACT 300
 DB 241 GAATCATCTCAACTCTAGTAAAGTCAAACTGTTTCTAAAGTGTATTTTAACT 300
 QY 301 ATTATATGCAATTTTGAACAGCTCTTTCACTTACTAGCAATTTTATCAGCAACACTT 360
 DB 301 ATTATATGCAATTTTGAACAGCTCTTTCACTTACTAGCAATTTTATCAGCAACACTT 360
 QY 361 GTATTGTCAGACTTTTAAAGTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 361 GTATTGTCAGACTTTTAAAGTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 421 NNN 480
 DB 421 NNN 480
 QY 481 NNN 540
 DB 481 NNN 540
 QY 541 NNN 600
 DB 541 NNN 600
 QY 601 NNN 660
 DB 601 NNN 660
 QY 661 ACCTGGTTAAATATTTTCAAAATCGCATATATATATATATATATATATATATATAT 720
 DB 661 ACCTGGTTAAATATTTTCAAAATCGCATATATATATATATATATATATATATATAT 720
 QY 721 TTTTGTCTTTTTCACATTTTGTCTTTTGTCTTTTGTCTTTTGTCTTTTGTCTTTTGTCT 780
 DB 721 TTTTGTCTTTTTCACATTTTGTCTTTTGTCTTTTGTCTTTTGTCTTTTGTCTTTTGTCT 780
 QY 781 TGAGATAAGTCTTTTTCATGTTTTCCTTATGAATAAATATATATATATATATATATAT 840
 DB 781 TGAGATAAGTCTTTTTCATGTTTTCCTTATGAATAAATATATATATATATATATATAT 840
 QY 841 GTAGCTGAGACCAATGATAGCTG 863

DB 841 GTAGCTGAGACCAATGATAGCTG 863
 RESULT 2
 ID ABK93491
 XX ABK93491 standard; cDNA; 392 BP.
 AC ABK93491;
 DT 23-AUG-2002 (first entry)
 XX Human breast specific nucleic acid, BSNA #97.
 DE Human; ss; breast specific nucleic acid; BSNA; breast cancer;
 KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
 XX Homo sapiens.
 OS WO200236807-A2.
 PN 10-MAY-2002.
 PD 29-OCT-2001; 2001WO-US046888.
 PF 27-OCT-2000; 2000US-0243802P.
 PR (DIAD-) DIADEXUS INC.
 PA Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
 PI WPI; 2002-463415/49.
 DR New breast-specific nucleic acids and polypeptides, useful for
 XX identifying, diagnosing, monitoring, staging, imaging, and treating
 PT breast cancer and non-cancerous disease states in breast tissues.
 PS Claim 1; Page 191; 281pp; English.

The invention relates to breast specific nucleic acids (BSNA) and breast-specific polypeptides (BSP). Also included are a method for determining the BSNA in a sample, a vector comprising a BSNA, a host cell comprising the vector, a method for producing a polypeptide encoded by a BSNA, an anti-BSP antibody and a method for determining the presence of a BSP in a sample. The breast-specific nucleic acids, polypeptides and compositions comprising them are useful for identifying, diagnosing, monitoring, staging, imaging, and treating breast cancer, mammary tumour and non-cancerous disease states in breast tissue; for identifying breast tissue; for monitoring, identifying and/or designing agonists and antagonists of the polypeptides; in gene therapy; in producing transgenic animals and cells; for producing engineered breast tissue for treatment and research; and as elements in an array or computer program for pattern recognition of breast disorders. The nucleic acids may be used as hybridisation probes to detect, characterise and quantify hybridising nucleic acids in, and isolate hybridising nucleic acids from, both genomic and transcript-derived nucleic acid samples. The BSP protein may be used in a vaccine composition for raising an immune response against breast cancer. The present sequence is BSNA cDNA of the invention

Query Match 45.4%; Score 392; DB 6; Length 392;
 Best Local Similarity 100.0%; Pred. No. 3.7e-72;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGGTCATTCAACTTTAAGTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 DB 1 ATGGGTTGGTCATTCAACTTTAAGTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 QY 61 TGGTTGATGATTTTACATTTATAGTATTTACATTCATGATGATTTTGTAGCTG 120
 DB 61 TGGTTGATGATTTTACATTTATAGTATTTACATTCATGATGATTTTGTAGCTG 120
 QY 121 TCCCTCCTATTGTTTAAACATTTGCGTCTCTCATTTTGGACCTATTTTAAACAGGCTCT 180

```

Db 121 TCCTCCTATTGTTTAAACATTTGGCTTCTCTCAATTTGACCTATTTTAAACAGGCTCT 180
Qy 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGCAATAAAAGTGGCT 240
Db 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGCAATAAAAGTGGCT 240
Qy 241 GAATCATCTTCAACTCTAGTAAGATGTCAACTGTTTCTAAAGTGTTTTATTATTAACT 300
Db 241 GAATCATCTTCAACTCTAGTAAGATGTCAACTGTTTCTAAAGTGTTTTATTATTAACT 300
Qy 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTTACTAGCAATTTATTATCAGCAACACTT 360
Db 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTTACTAGCAATTTATTATCAGCAACACTT 360
Qy 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392
Db 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392

RESULT 3
AAL26040/c
ID AAL26040 standard; cDNA; 229 BP.
XX AC AAL26040;
XX XX
XX 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 18497.
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX OS
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US0000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR
XX PT New peptide useful as a marker for the diagnosis of breast cancer.
XX PS Claim 1; Page 3412; 3695pp; English.
XX SQ
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity
XX Sequence 229 BP; 60 A; 55 C; 51 G; 63 T; 0 U; 0 Other;
XX Query Match 10.4%; Score 89.8; DB 4; Length 229;
XX Best Local Similarity 97.8%; Pred. No. 1.6e-09;
XX Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 121 TCCTCCTATTGTTTAAACATTTGGCTTCTCTCAATTTGACCTATTTTAAACAGGCTCT 180
Qy 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGCAATAAAAGTGGCT 240
Db 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGCAATAAAAGTGGCT 240
Qy 241 GAATCATCTTCAACTCTAGTAAGATGTCAACTGTTTCTAAAGTGTTTTATTATTAACT 300
Db 241 GAATCATCTTCAACTCTAGTAAGATGTCAACTGTTTCTAAAGTGTTTTATTATTAACT 300
Qy 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTTACTAGCAATTTATTATCAGCAACACTT 360
Db 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTTACTAGCAATTTATTATCAGCAACACTT 360
Qy 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392
Db 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392

RESULT 3
AAL26040/c
ID AAL26040 standard; cDNA; 229 BP.
XX AC AAL26040;
XX XX
XX 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 18497.
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX OS
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US0000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR
XX PT New peptide useful as a marker for the diagnosis of breast cancer.
XX PS Claim 1; Page 3412; 3695pp; English.
XX SQ
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity
XX Sequence 229 BP; 60 A; 55 C; 51 G; 63 T; 0 U; 0 Other;
XX Query Match 10.4%; Score 89.8; DB 4; Length 229;
XX Best Local Similarity 97.8%; Pred. No. 1.6e-09;
XX Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 169 AAACAGGCTCTTTAAATATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 170 AAAATGGCTCTTTAAATATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 111
Qy 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 110 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 78

RESULT 4
AAL16838/c
ID AAL16838 standard; cDNA; 233 BP.
XX AC AAL16838;
XX XX
XX 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 9295.
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX OS
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US0000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR
XX PT New peptide useful as a marker for the diagnosis of breast cancer.
XX PS Claim 1; Page 1670; 3695pp; English.
XX SQ
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity
XX Sequence 233 BP; 61 A; 55 C; 53 G; 64 T; 0 U; 0 Other;
XX Query Match 10.4%; Score 89.8; DB 4; Length 233;
XX Best Local Similarity 97.8%; Pred. No. 1.6e-09;
XX Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 169 AAACAGGCTCTTTAAATATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 174 AAAATGGCTCTTTAAATATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 115
Qy 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 114 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 82

```

```

RESULT 5
AAL17847/c
ID AAL17847 standard; cDNA, 256 BP.
XX
AC AAL17847;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 10304.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 1837; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 256 BP; 72 A; 56 C; 57 G; 71 T; 0 U; 0 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 256;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
DB 174 AAAATGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 115
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 114 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 82
RESULT 6
AAL17770
ID AAL17770 standard; cDNA; 258 BP.
XX
AC AAL17770;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 10227.
XX

```

```

KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 1823; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 258 BP; 70 A; 59 C; 57 G; 72 T; 0 U; 0 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 258;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
DB 85 AAAATGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 144
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 145 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 177
RESULT 7
AAL10695
ID AAL10695 standard; cDNA; 270 BP.
XX
AC AAL10695;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 3152.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX

```

PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
DR
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX
PS Claim 1; Page 583; 3695pp; English.
XX
XX
CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity
XX
SQ Sequence 270 BP; 70 A; 61 C; 63 G; 73 T; 0 U; 3 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 270;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
DB 97 AAAATGGCCCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 156
QY 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 157 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 189
RESULT 8
AAL10772/c
ID AAL10772 standard; cDNA; 298 BP.
XX
XX
AC AAL10772;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 3229.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX

PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
DR
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX
PS Claim 1; Page 596; 3695pp; English.
XX
XX
CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity
XX
SQ Sequence 298 BP; 77 A; 69 C; 73 G; 78 T; 0 U; 1 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 298;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
DB 213 AAAATGGCCCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 154
QY 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 153 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 121
RESULT 9
AAL17429
ID AAL17429 standard; cDNA; 317 BP.
XX
XX
AC AAL17429;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 9886.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US000798.
XX
XX 14-JAN-2000; 2000US-0176077P.
XX 14-MAR-2000; 2000US-0189167P.
XX 24-MAR-2000; 2000US-0192099P.
XX 29-MAR-2000; 2000US-0193480P.
XX 15-MAY-2000; 2000US-0205230P.
XX 09-JUN-2000; 2000US-0211315P.
XX 25-JUL-2000; 2000US-0220534P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
DR
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX
PS Claim 1; Page 1762; 3695pp; English.
XX
XX
CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is

Sequence 332 BP; 86 A; 75 C; 85 G; 86 T; 0 U; 0 Other; ;

Query Match	9.8%	Score 85	DB 4	Length 525
Best Local Similarity	94.6%	Prod. NO. 1.8e-08		
Matches	88	Conservative	0	Mismatches 0; Indels 0; Gaps 0;

QY	169	AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTPAGAAATGGAGTGCCTGGA	228
Db	387	AAAATGGCCTTTTAAATATACACTTCTCTGTAGTGTATGCTPAGAAATGGAGTGCCTGGA	328

Qy 229 ATAAAGTGGTGAATCATCTTCACTCTAGTA 261

Db 327 ATAAAGTGGCTGATCATCTCAACTTAGTA 295

RESULT 12
AAL08295/c
ID AAL08295 standard; cDNA; 248 BP.
XX
XX AC AAL08295;
XX
XX DT 07-DEC-2001 (first entry)
XX
XX DE Human breast cancer expressed polynucleotide 752.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX
XX PF 10-JAN-2001; 2001WO-US000798.
XX
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
XX DR New peptide useful as a marker for the diagnosis of breast cancer.
XX PT
XX PS Claim 1; Page 208; 3695pp; English.
XX
XX CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity

XX SQ Sequence 248 BP; 61 A; 61 C; 61 G; 65 T; 0 U; 0 Other;
Query Match 9.0%; Score 77.8; DB 4; Length 248;
Best Local Similarity 96.8%; Pred. No. 5.1e-07;
Matches 90; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 169 AAACAGGCTCTTTAAATATACACTTCTCTGATGATGCTAGAAATGAGTGGCTGA 228
DB 189 AAAATGGCTCTTTAAATATACACTTCTCTGATGATGCTAGAAATGAGTGGCTGA 131

QY 229 ATAAAGTGGCTGAATCATCTTCAACTTAGTA 261
DB 130 ATAAAGTGGCTGAATCATCTTCAACTTAGTA 98

RESULT 13
AAL13109/c
ID AAL13109 standard; cDNA; 534 BP.
XX
XX AC AAL13109;
XX
XX DT 07-DEC-2001 (first entry)

XX
XX DE Human breast cancer expressed polynucleotide 5566.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX
XX PF 10-JAN-2001; 2001WO-US000798.
XX
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
XX DR New peptide useful as a marker for the diagnosis of breast cancer.
XX PT
XX PS Claim 1; Page 1000-1001; 3695pp; English.
XX
XX CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity

XX SQ Sequence 534 BP; 119 A; 121 C; 119 G; 142 T; 0 U; 33 Other;
Query Match 8.3%; Score 71.6; DB 4; Length 534;
Best Local Similarity 90.4%; Pred. No. 1.1e-05;
Matches 85; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 169 AAACAGGCTCTTTAAATATACACTTCTCTGATGATGCTAGAAATGAGTGGCTGG 227
DB 265 AAAATGGCTCTTTAAANANACACTTCTCTGATGATGCTAGAAATGAGTGGCTGG 206

QY 228 ATAAAGTGGCTGAATCATCTTCAACTTAGTA 261
DB 205 ATAAAGTGGCTGAATCATCTTCAACTTAGTA 172

RESULT 14
AAS61094
ID AAS61094 standard; DNA; 6062 BP.
XX
XX AC AAS61094;
XX
XX DT 29-JAN-2002 (first entry)
XX
XX DE Human gene regulation-associated gene oligonucleotide #49.
XX
XX KW Human; Gene regulation-associated gene; severe combined immunodeficiency; cardiac damage; inflammatory response; Haemophilia; Werner syndrome; asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome; renal disease; Preeclampsia; cardiac allograft vascular disease; colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour; immunostimulant; cardiac; antiinflammatory; coagulant; antithrombotic; nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

XX Homo sapiens.
 OS WO200177375-A2.
 PN 18-OCT-2001.
 XX
 PD
 XX
 PF
 XX
 XX
 XX
 PR 06-APR-2001; 2001WO-EP003968.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-017470/02.
 XX
 XX New nucleic acid sequences from chemically modified genes associated with
 PT gene regulation, useful for analyzing cytosine methylations for diagnosis
 PT and therapy of diseases e.g. severe combined immunodeficiency disease.
 XX
 XX Claim 1; SEQ ID NO 50; 26pp; English.
 XX
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The chemical
 CC pretreatment converts cytosine bases unmethylated at the 5-position to
 CC uracil or another base with hybridisation behaviour dissimilar to
 CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
 CC oligomers (or sets/arrays) and method are useful in the diagnosis of
 CC diseases (or predisposition to diseases) associated with gene regulation
 CC and in therapy of such diseases, by enabling analysis of the cytosine
 CC methylation patterns of such genes, kits are provided. They are
 CC especially useful in diagnosis and therapy of e.g. severe combined
 CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
 CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
 CC syndrome, renal disease, preclampsia, graft versus-host disease. The
 CC present sequence is a sequence included in the sequence data for this
 CC specification and is associated with the human gene regulation-associated
 CC genes. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
 XX
 XX Sequence 6062 BP; 1555 A; 215 C; 1265 G; 3027 T; 0 U; 0 Other;
 SQ
 Query Match 7.8%; Score 67.6; DB 6; Length 6062;
 Best Local Similarity 32.8%; Pred. No. 9.6e-05;
 Matches 270; Conservative 0; Mismatches 552; Indels 1; Gaps 1;
 20 TTTAAGTTTATGAGATATCCATGTTGAATTTTGTAGCTGCTGGTTGATGATTTTACA 79
 5122 TTTAGTCGCTGATATATATAGTAATTTTAAATGAGTTTGTAGTTTGTATTTT 5181
 80 TTATATAGTATTACATCCATGATAGTCTTCAGTAGATAATCCTCTATTGTTTAAACAT 139
 5182 TTGATTATAAATTATATGTTTATTGTTTAAATATTGTTTGGATGTTTAAAGAT 5241
 140 TTGGCTGCTCTCATTTGACCTATTTTTAAACAGGCTCTTTAAATATACACTTCTCTG 199
 5242 ATTTTAAAAATTTTAAATTTAAATTTTATTATTATTGTTTATTGTTTAAATTTTAA 5301
 200 TAGTGTATGCTAGAAATGGAGTGGCTGCAATAAAGTGGCTGAATCATCTTCAACTCTAG 259
 5302 AAGTGGGTATTTAAT-TGCTTTTGTAAATAGAAATAGGGTATTATTATTAGTCTG 5360
 260 TAAGATGCAAACTGTTTCTAAAGTGTGTTTATTTTAACTATATATATGCAATTTGAA 319
 5361 TTTTGTGTTAGTATTAGAAAGTGTGTTTATTTAGTTTGTGTTTAAATTTTAAATTTTA 5420
 320 CAGCTCTTCACTTACTAGCAATTTATATCAGCAACACTTGTATTGTCAGACTTTTAA 5420

Db 5421 TTTTTCGTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 5480
 QY 380 GTTTTCATTCTACGNN 439
 Db 5481 GTAGTAGTGTGTTTTCGTCGGTAAATATGTAATTTATTTTATATCGTATTTAGAGTAT 5540
 QY 440 NNN 499
 Db 5541 TTTTTCATAATTTAATTTATGTTAGTTTTCGTTTAAAGTTTTCGTTTAAAGTTTTC 5600
 QY 500 NNN 559
 Db 5601 ATTTGATGATATATTTTATTTTTCGTTAAATTTGTTATATTTTGTATTTTAAAT 5660
 QY 560 NNN 619
 Db 5661 TTAGTAAGTTTGTGTTGTTTATTTATTTTATTTTATTTTATTTTATTTTATTTT 5720
 QY 620 NNNNNNGTGATCATATTATGTTTTCCTCATTTAAACAAACCTGGTAAATATTTTAC 679
 Db 5721 TTTTTCGTTTAAAGTTTTCGTTTAAATTTATTTTTCGTTTAAAGTTTTCGTTTAA 5780
 QY 680 AAATCGACATTATGATATATATCTCCAAATTTTAAATTTTGTCTTTTTCACATT 739
 Db 5781 AATTAGGTAGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5840
 QY 740 TTAGTCTTTAGCTCAGCTGGAATTCATTTCTGCTGGTGGTGGAGATAAGTCTTTTCA 799
 Db 5841 TAGTAGTAGAGTTTGTAAATATTTGTTTATTTTAAAGAGATAAATTTAAGTTTAA 5900
 QY 800 GTTTTCCCTATGAAATTAATTTTCTCTCTGTTGCGAGT 842
 Db 5901 TTTTTCGTTTAAAGTTTTCGTTTATTTTATTTTATTTTATTTTATTTTATTTT 5943
 RESULT 15
 ID ABL32591 standard; DNA; 6195 BP.
 AC ABL32591;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 564.
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.
 OS Homo sapiens.
 PN WO200200928-A2.
 PD 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for

PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

XX Claim 1; SEQ ID NO 564; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention

XX
SQ Sequence 6195 BP; 1636 A; 57 C; 1209 G; 3293 T; 0 U; 0 Other;

Query Match 7.7%; Score 66.2; DB 6; Length 6195;

Best Local Similarity 34.2%; Pred. No. 0.00019;

Matches 269; Conservative 0; Mismatches 511; Indels 6; Gaps 3;

Qy 21 TTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGGTTTGATGATTTTACAT 80

Db 3391 TTATATTTTATTTTATTTATTTATTTATTTAGTAATATTTAGGTTAATTTATATTTGTT 3450

Qy 81 TATATAGTATACATT-CCATGATAGTCTCAGTAGATAATCTCCTATTTGTTTACAT 139

Db 3451 TATTTTGAATAAATGTTGAATGGAATGCGAGGTAGATTTTTTTGTATATAGTGATTT 3510

Qy 140 TTGGTTTGTCTCTCATTTTGACCTATTTTAAACAGGCTCTTTAAATATACACTTCTCTG 199

Db 3511 TAATTTTGGGAGATATTTAGAAAGTGGATTTGGGTATATGGTAAATTTATGTT 3570

Qy 200 TAGTGATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTGAATCATCTCACTCTAG 259

Db 3571 TAGGTTTGGAGGAATCGTTATATTTTAAATGGTTGTATTAATTTATATTTTAA 3630

Qy 260 TAAGATGCAACTGTTTCTAAAGTGTTTTATTTAACTATTTATATGCAATTTGAA 319

Db 3631 TTAAGGTTTAAAGGTTTATTTTATATTTTATTTTATTTTAA--ATACGATTTATTTGTT 3688

Qy 320 CAGCTCTTTCACTTACTAGCAATTTATTCAGCAACACTTGTATTGTCAGACTTTAA 379

Db 3689 TATTTGTTAAAGTTATTTTAAATAGATGATTTATTTTAAATTTGTTATTTT 3748

Qy 380 GTTTTCATTCACGNN 439

Db 3749 TTGATAATTAGTATGTCGAGTGTCTTTTATATATTTTGTGGGTAGGTTTATGTTT 3808

Qy 440 NNN 499

Db 3809 TTTTAAATTCAGATGTTTGTCTTGTATTCAGTGTGTTGAATTTTATATTTTGG 3868

Qy 500 NNN 559

Db 3869 GATATTAGTTTTTATTAGATATATGTTTGTAAATATTTTATTTTATTTTGGATTAT 3928

Qy 560 NNN 619

Db 3929 TTTTGTGTTTTTTTGTGTAGAGTTTTTTAGTGTATATTTTGTGTTTTTGGTTTTT 3988

Qy 620 NNNNNNNNGTATCATATTTATGTTTGTCTCAATTTAAACCTGTTAAATATTTTAC 679

Db 3989 TGAATTTGTTGTTTTTGTGGAGTTTTGTTGAGAAATAATTTAGAAAAATGTT 4048

Qy 680 AAATCGACATTATGATATATATCTCCAAATTTTAAATTTTGTCTTTTTCACATT 739

Db 4049 ATAAAGATTTTTTTTATGTTTTTTTATGATTTTATATTTTGGGTTTTTT--ATG 4105

Qy 740 TTAGTCTTTAGCTCAGCTGGAAATCAATTTCTGCTGTGCTGAGATAAGTCTTTTTCAT 799

Db 4106 TAAGTTTTTATTTATTTGATTTGAGTTTGTGTATAGTGTGGATAAGTGTATTT 4165

Qy 800 GTTTTT 805

Db 4166 TATTTT 4171

Search completed: August 14, 2004, 19:13:42
Job time : 437.526 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 19:13:54 ; Search time 490.203 Seconds
(without alignments)
8638.053 Million cell updates/sec

Title: US-10-082-830-98

Perfect score: 863
Sequence: 1 atgggttggtcattcaact.....gctgagaccaatgatagctg 863

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/pct_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	630	73.0	863	15	US-10-082-830-98
2	392	45.4	392	15	Sequence 98, Appl
3	91.2	10.6	1017	15	Sequence 97, Appl
4	89.8	10.4	390	15	Sequence 11216, A
5	89.8	10.4	393	15	Sequence 10053, A
6	89.8	10.4	441	15	Sequence 887, App
7	89.8	10.4	481	15	Sequence 2119, Ap
8	71.6	8.3	1015	15	Sequence 2037, Ap
9	67.6	7.8	6062	13	Sequence 4377, Ap
10	66.2	7.7	6195	15	Sequence 50, Appl
11	64.4	7.5	1214	13	Sequence 564, Appl
12	63.6	7.4	3673778	15	Sequence 102083,
13	62.8	7.3	7128	15	Sequence 1, Appli
14	62.8	7.3	40324	17	Sequence 1532, Ap
					Sequence 180, App

15	60.6	7.0	9964	15	US-10-311-455-71	Sequence 71, Appl
16	60.4	7.0	14147	13	US-10-221-714A-470	Sequence 470, App
17	60.4	7.0	14147	15	US-10-172-086-52	Sequence 52, Appl
18	60.4	7.0	14147	17	US-10-311-507-40	Sequence 40, Appl
19	59.6	6.9	7442	13	US-10-221-714A-409	Sequence 409, Appl
20	59.6	6.9	15732	15	US-10-239-676-95	Sequence 95, Appl
21	59.6	6.9	15732	15	US-10-240-453-107	Sequence 107, App
22	59.4	6.9	6171	15	US-10-311-455-983	Sequence 983, App
23	58.8	6.8	5327	15	US-10-239-676-61	Sequence 61, Appl
24	58.8	6.8	5327	15	US-10-240-453-63	Sequence 63, Appl
25	58.8	6.8	9539	15	US-10-239-676-52	Sequence 52, Appl
26	58.8	6.8	9539	15	US-10-240-453-54	Sequence 54, Appl
27	58.6	6.8	12237	15	US-10-311-455-2331	Sequence 2331, Ap
28	58.4	6.8	3673778	15	US-10-312-841-2	Sequence 2, Appli
29	58.2	6.7	6132	15	US-10-311-455-836	Sequence 836, App
30	58	6.7	12356	13	US-10-221-714A-232	Sequence 232, App
31	57.6	6.7	6000	15	US-10-311-455-2278	Sequence 2278, Ap
32	57.6	6.7	6292	13	US-10-221-714A-461	Sequence 461, App
33	57.6	6.7	10467	15	US-10-204-708-2	Sequence 2, Appli
34	57.2	6.6	5362	15	US-10-240-453-287	Sequence 287, App
35	57	6.6	6668	15	US-10-311-455-1670	Sequence 1670, Ap
36	57	6.6	12138	15	US-10-311-455-1602	Sequence 1602, App
37	57	6.6	12138	17	US-10-257-166-116	Sequence 116, App
38	57	6.6	18283	13	US-10-221-613-325	Sequence 325, App
39	55.8	6.5	523	13	US-10-027-632-195935	Sequence 195935,
40	55.8	6.5	523	16	US-10-027-632-195935	Sequence 195935,
41	55.8	6.5	9238	15	US-10-240-453-240	Sequence 240, App
42	55.8	6.5	15387	15	US-10-311-455-157	Sequence 157, App
43	55.6	6.4	1300	17	US-10-668-749A-1	Sequence 1, Appli
44	55.6	6.4	6175	15	US-10-311-455-1280	Sequence 1280, Ap
45	55.6	6.4	7544	15	US-10-239-676-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-10-082-830-98
; Sequence 98, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatenIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 863
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (395)..(627)
; OTHER INFORMATION: a, c, g or t
US-10-082-830-98

Query Match 73.0%; Score 630; DB 15; Length 863;

Best Local Similarity 100.0%; Pred. No. 3.9e-119;

Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGGTCATTCACTTTAACTTTATGAGATATATCCATGTGAATTTGTAGCTG 60

Db 1 ATGGGTGGTCATTCACTTTAACTTTATGAGATATATCCATGTGAATTTGTAGCTG 60

QY 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
Db 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
QY 121 TCCTCCATGTTTAAACATTTGCGTTCCTCTCATTTGACCTATTTTAAACAGGCTCT 180
Db 121 TCCTCCATGTTTAAACATTTGCGTTCCTCTCATTTGACCTATTTTAAACAGGCTCT 180
QY 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
Db 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
QY 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTAAAGTGTATTTTAACT 300
Db 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTAAAGTGTATTTTAACT 300
QY 301 ATTATATGCTCAATTTGAACAGCTCTTCACTTACTAGCAATTTATATCAGCAACACTT 360
Db 301 ATTATATGCTCAATTTGAACAGCTCTTCACTTACTAGCAATTTATATCAGCAACACTT 360
QY 361 GTTATGTGACACTTTTAAAGTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 420
Db 361 GTTATGTGACACTTTTAAAGTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 420
QY 421 NNN 480
Db 421 NNN 480
QY 481 NNN 540
Db 481 NNN 540
QY 541 NNN 600
Db 541 NNN 600
QY 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGATCATATTTATGTTTGCTCATTTAAATA 660
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGATCATATTTATGTTTGCTCATTTAAATA 660
QY 661 ACCTGGTTAAATATTTACAAATCGACATTATGATATTTATCTTCCAAATTTTAAATA 720
Db 661 ACCTGGTTAAATATTTACAAATCGACATTATGATATTTATCTTCCAAATTTTAAATA 720
QY 721 TTTTGTCTTTTTCACATTTTAGTCTTACGCTGGAATTCATTTCTGTGTGGTG 780
Db 721 TTTTGTCTTTTTCACATTTTAGTCTTACGCTGGAATTCATTTCTGTGTGGTG 780
QY 781 TGAGATAAGTCTTTTTCATGTTTTCCTATGAAATAAATATTTCTTCTGTATTCGAG 840
Db 781 TGAGATAAGTCTTTTTCATGTTTTCCTATGAAATAAATATTTCTTCTGTATTCGAG 840
QY 841 GTAGCTGAGCAATGATAGCTG 863
Db 841 GTAGCTGAGCAATGATAGCTG 863

RESULT 2

US-10-082-830-97
; Sequence 97, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; PRIORITY FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802

QY 1 ATGGGTTTGGTCATTCACATTTAAAGTTTATGAGATATATCCATGTTGCAATTTTGTAGCTG 60
Db 1 ATGGGTTTGGTCATTCACATTTAAAGTTTATGAGATATATCCATGTTGCAATTTTGTAGCTG 60
QY 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
Db 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
QY 121 TCCTCCATGTTTAAACATTTGCGTTCCTCTCATTTGACCTATTTTAAACAGGCTCT 180
Db 121 TCCTCCATGTTTAAACATTTGCGTTCCTCTCATTTGACCTATTTTAAACAGGCTCT 180
QY 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
Db 181 TTAATATACACTTCTCTGTAGTGTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
QY 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTAAAGTGTATTTTAACT 300
Db 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTAAAGTGTATTTTAACT 300
QY 301 ATTATATGCTCAATTTGAACAGCTCTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 360
Db 301 ATTATATGCTCAATTTGAACAGCTCTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 360
QY 361 GTTATGTGACACTTTTAAAGTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 392
Db 361 GTTATGTGACACTTTTAAAGTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNN 392

RESULT 3

US-10-198-846-11216
; Sequence 11216, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ-ID NO 11216
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1017
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-11216
Query Match 10.6%; Score 91.2; DB 15; Length 1017;
Best Local Similarity 96.9%; Pred. No. 5.9e-09;

```
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 166 TTTAAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCT 225
Db 367 TCTGAAAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCT 426
QY 226 GGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 427 GGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 462

RESULT 4
US-10-198-846-10053
; Sequence 10053, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10053
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 386, 387, 388, 389, 390
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10053

Query Match 10.4%; Score 89.8; DB 15; Length 390;
Best Local Similarity 97.8%; Pred. No. 7.9e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
Db 127 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 186
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 187 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 219

RESULT 5
US-10-198-846-887
; Sequence 887, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887
```

```
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 393
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-887

Query Match 10.4%; Score 89.8; DB 15; Length 393;
Best Local Similarity 97.8%; Pred. No. 8e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
Db 178 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 237
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 238 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 270

RESULT 6
US-10-198-846-2119/c
; Sequence 2119, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2119
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3, 53
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2119

Query Match 10.4%; Score 89.8; DB 15; Length 441;
Best Local Similarity 97.8%; Pred. No. 8.3e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
Db 216 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 157
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 156 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 124

RESULT 7
US-10-198-846-2043
; Sequence 2043, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
```

```

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2043
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 2, 3, 4, 5, 13, 14, 16, 17, 21, 27, 37, 48, 74, 93, 101,
; LOCATION: 333, 464, 478
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2043

Query Match      10.4%; Score 89.8; DB 15; Length 481;
Best Local Similarity 97.8%; Pred. No. 8.6e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 169 AAACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 151 AAAATGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 210

Qy 229 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 261
Db 211 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 243

RESULT 8
US-10-198-846-4377/c
; Sequence 4377, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4377
; LENGTH: 1015
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 5, 7, 11, 14, 15, 16, 18, 19, 20, 22, 23, 27, 29, 31, 32,
; LOCATION: 33, 34, 35, 39, 41, 43, 49, 50, 53, 54, 55, 56, 57, 59,
; LOCATION: 61, 62, 63, 65, 67, 69, 73, 75, 81, 82, 83, 84, 85, 86,
; LOCATION: 87, 89, 90, 93, 94, 96, 97, 98, 99, 103, 106, 108, 109,
; LOCATION: 110
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 111, 113, 121, 124, 130, 139, 143, 155, 175, 186, 188, 191,
; LOCATION: 193, 212, 256, 282, 298, 308, 334, 374, 399, 401, 442, 459,
; LOCATION: 462, 472, 495, 498, 502, 512, 515, 516, 569, 585, 602,
; LOCATION: 618, 641, 655, 656, 695, 715, 731, 744, 755, 758, 775
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: 808, 809, 810, 842, 847, 854, 866, 880, 905, 907, 908, 921,
; LOCATION: 924, 925, 926, 933, 934, 937, 939, 940, 951, 957, 958, 963,
; LOCATION: 969, 994, 996, 1013
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-4377

Query Match      8.3%; Score 71.6; DB 15; Length 1015;
Best Local Similarity 90.4%; Pred. No. 6e-05;
Matches 85; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Qy 169 AAACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGG 227
Db 418 AAAATGGCGCTTTAAANANACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGG 359

Qy 228 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 261
Db 358 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 325

RESULT 9
US-10-221-613-50
; Sequence 50, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221.613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 50
; LENGTH: 6062
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-50

Query Match      7.8%; Score 67.6; DB 13; Length 6062;
Best Local Similarity 32.8%; Pred. No. 0.00079;
Matches 270; Conservative 0; Mismatches 552; Indels 1; Gaps 1;

Qy 20 TTAAAGTTTATGAGATATATCCATGTTGAAATTTTGTAGCTGTGGTTGATGATTTTACA 79
Db 5122 TTTAGTCGCTATATATATTAGTAAATTTTAAATGTAAGTTTGTAGTTAGTTTGT 5181

Qy 80 TTATATAGTATATACATCCATGGATAGTCTCTCAGTAGATAATCCTCTATGTTTAAACAT 139
Db 5182 TTGATATAAATTTATATAGTTTATGTTTAAAAATATTGTTTTCGATGTTTAAAGAT 5241

Qy 140 TTGGCTTCTCTCTCATTTTGACCTATTTTTAAACAGGCGCTTTTAAATATACACTTCTG 199
Db 5242 ATTTTAAAAATTTTAAATTTAAATTTTATTTTATTAGTTTATTTTAAATTTTAA 5301

Qy 200 TAGTGTATGCTAGAAATGGAGTGGCTGGAATAAAGTGGCTGAATCATCTTCAACTCTAG 259
Db 5302 AAGTGGGTATTATTAAT-TCGTTTTTGTAAATAGAAATATGGGTATTTATTAGTTCGT 5360

```

260 TAAGATGTCACAACTGTTTCTAAAGAGTGTATTTTAACTATATATGCAATTTGAA 319
3361 TTTTITTTTGTAGTATTAGAAAGTTTATAGTGTATTTTATTTTAAATTTTAA 5420
320 CAGCTCTTTCACCTACTAGCAATTTATATATCAGCAACACTGTTTATGTCAGACTTTAA 379
5421 TTTTITTTTCTGTTTATTTATTTATTTATTTTAAAGTATATAATTTTGTGTTTAAATTTT 5480
380 GTTTCATTCACCTGCGNN 439
5481 GTAGTAGTGTGTTTGTGCGGTAATATGTAATTTATTTTATATCGTATTAGAGTGAT 5540
440 NNN 499
5541 TTTTITTTAAATTTTAAATTTATGTTAGTGTGTTTGTGTTTAAAGTGTGTTTAAAGAA 5600
500 NNN 559
5601 ATTGTGATGATATATTTTATTTTATTTTGGTAAATTTGTTTATTTGTTTATTTTAAAT 5660
560 NNN 619
5661 TTAGTAGTGTGTTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5720
620 NNNNNNNNGTATCATATTTATGTTTGTCTCATTTAAAGAACTGTTTAAATTTTAC 679
5721 TTTTITTTTAAAGTGTGTTTAAATTTATTTTATTTTAAAGTGTGTTTATTTGATTTAA 5780
680 AAATCGACATTTATGATATATTTCTCCAAATTTTAAATTTTCTGTTTCTTTTTCACAT 739
5781 AATTAGGTTAGGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5840
740 TTAGTCTTTAGCTCAGCTGGAATTCATTTCTGTTGTTGTTGTTGTTGTTGTTGTTT 799
5841 TAGTAGTAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5900
800 GTTTCCTTCATGAAATTAATTTCTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 842
5901 TTTTITTTAAAGTTAAGATTTTGTGTTTATTTTGTGTTTATTTTGTGTTTATTTTGT 5943

RESULT 10

US-10-311-455-564
; Sequence 564, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation Status of Cytosine Residues in DNA
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043926.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 564
; LENGTH: 6195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-564

Query Match 7.7%; Score 66.2; DB 15; Length 6195;
Best Local Similarity 34.2%; Pred. No. 0.0015;
Matches 269; Conservative 0; Mismatches 511; Indels 6; Gaps 3;

21 TTAAGTTTATGAGATATATCCATGTTGAAATTTTGTAGTGTGGTTTGATGATTTTACAT 80
3391 TTAATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3450
81 TATATAGTATTACATTTCCATGATAGTCTCTCAGTAGATAATCTCTCTCTCTCTCTCT 139
3451 TATTTTGAATATGTTGAATGATATGGAGTGTAGATTTTCTTTTGTATATAGTATTT 3510
140 TGGCTTCT 199
3511 TAAATTTTGGGAGATATATTTAGAGTGGGATTTGTTGGGTTATATGTAATTTATGTT 3570
200 TAGTGTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCTGGAATCATCTTCAACTCTAG 259
3571 TAGTGTGTTGAGGAATCGTTATATTTTAAATGAGTGTATTTTATTTATTTATTTT 3630
260 TAAGATGTCACAACTGTTTCTTAAAGTGTGTTTATTTTAACTATTATATGTCATTTTCAA 319
3631 TTAATGTTTATAGGGTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3688
320 CAGCTCTTCT 379
3689 TATTTGTTTAAAGTGTATTTTAAATGATGATTTATTTTATGATTTTAAATTTGATTTT 3748
380 GTTTCATTCACCTCGNN 439
3749 TTGATAATTAGTAGTGTGCTGAGTGTGTTTATATATTTTGTGGTGTAGTGTATTTGTTT 3808
440 NNN 499
3809 TTTTITTTAAATTTGAGATGTTGTTTCTGTTTATTTGAGTGTGTTTAAATTTTATTTT 3868
500 NNN 559
3869 GATATTAGTGTGTTTATTTAGATATATGTTTCTGTTGTTGTTGTTGTTGTTGTTGTT 3928
560 NNN 619
3929 TTTTITTTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3988
620 NNNNNNNNGTATCATATTTATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 679
3989 TGATTTTCTGTTTGTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4048
680 AAATCGACATATGATATATTTATTTCTCCAAATTTTAAATTTTATTTGTTCTTTTTCACAT 739
4049 ATAAAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4105
740 TTAGTCTTTAGCTCAGCTGGAATTCATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 799
4106 TAAAGTTTAAATTTATTTTGTGTTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4165
800 GTTITTT 805
4166 TATTTT 4171

RESULT 11

US-10-424-599-102083/c
; Sequence 102083, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 102083


```
Db 38536 AGTTTTCGTAGATTAAAGAGTCTTTTATGAAGTCTTTTGTGTTGTTTTT 38595
Qy 661 ACCGTGTTAAATATTCACAAATCGACATTATGATATATCTCCAAAATTTAATAA 720
Db 38596 TATTTTATATGTTGTTTATTAATGAAAATCGAATATGTAATATGTAAGA 38655
Qy 721 TTTTGTCTTTTTCACATTTTAGCTCAGCTGGAATTCATTTCTGTGTGGTG 780
Db 38656 TTTTGGAGTTTATCGGGTTTTTTTTTTGAGTATTAATTTATGTTGTTGTTTG 38715
Qy 781 TGAGATAAGTCTTTTCAAGTTTTTCCCTATGAATAAATTAATTCCTTCGTAT 835
Db 38716 TTTGATAGTTAAATTTGTCGGAATTAATGTTAAATATGTTTATAAGGGTAT 38770

RESULT 15
US-10-311-455-71
; Sequence 71, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 71
; LENGTH: 9964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

Query Match 7.0%; Score 60.6; DB 15; Length 9964;
Best Local Similarity 33.2%; Pred. No. 0.026;
Matches 275; Conservative 0; Mismatches 547; Indels 6; Gaps 2;

Qy 2 TGGGTTTGGTCATTCAACTTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGT 61
Db 8315 TGTGTTTTAAATATAAGTTGAGGGTGTAGTAGATTTTATTATTAGTAATTTTTTTATA 8374
Qy 62 GGTGTTGATGTTTTTACATATATAGTATTACATTCATGATGATGTCAGTAGATAAT 121
Db 8375 GTGTAGTGTGTTTGTGTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 8434
Qy 122 CCTCCTATGTTTAAACATTTTGGTGTCTCTCTATTTTGACCTATTTTAAACAGGCTCTT 181
Db 8435 TTGTTTTTATTTTGTATTTGTTTTTTTAAATTTGTTTTTTTTTTTATTTTATTTGTTTTT 8494
Qy 182 TAAATATACATCTCTGTAGTGTAGCTAGAAATGGAGTGGCTGGAATAAAGTGGCTG 241
Db 8495 TTTTGTGTTTTTTTGTAGTGTGTTTTTTTGGTGTGTTTTTTAGTTTTTTTATTTTAAAT 8554
Qy 242 AATCATCTTCAACTCTAGTAAGATGTCAAAGCTGTTTCTAAAAGTGTGTTTTTAACTA 301
Db 8555 TTATTTTTTTTGTGTTAGTATTTTTTTGTTTTTTTATTTTGTGTTGTTTTTTTATTTTAA 8614
Qy 302 TTATATGTCATTTTGAACAGCTCTTTCACTTACTAGCAATTTTATATCAGCAACACTG 361
Db 8615 TTTTTTTTTTTTTTGTTTTTTAGTTTTTTTTTTTATTTT--TTTAATTAGGGTTATGATTT 8672
Qy 362 TTATGTGCAGACTTTTAAAGTTTTTCATTCACCTGGNNNNNNNNNNNNNNNNNNNNNN 421
```

```
Db 8673 TTTTAAAAATTTATTTAGTTTTTTTGTAAATTTTGGTAGTATTTTATGTTTGTGTT 8732
Qy 422 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 481
Db 8733 TGATGATTTTATTTTGTGTTGGAATTTTATTTATTTTGTGTTTATTTATGAAATTT 8792
Qy 482 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 541
Db 8793 TCFATAAATTTTTCAGATTTATTTGATTTTTTTTATGTTTTTTTCGTGTTTTTGTGTTAT 8852
Qy 542 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 601
Db 8853 TTGTTTTGTTTTTATTTTGTGTTGATTAATTAATTAATTTTATTTTATTTTATTTT 8912
Qy 602 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 661
Db 8913 TTTTTCGATTTAAATTTTGTGTTTTTATTTTATTTTATTTTATTTTATTTTATTTT 8972
Qy 662 CCTGTTAAATTTTCAAAATCGACATTATGATATATATCTTCCAAAATTTTAAATAT 721
Db 8973 GATTTTTTTTTTTTTTAAATTTTGTGTTTTTAAATTTAT---TATTAATTAATTAAT 9028
Qy 722 TTTGTCTTTTTTCACATTTTATGCTTTAGCTCAGCTGGAATTCATTTCTGTGTGGTGT 781
Db 9029 TTTGTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTAAATTTTAGATTTTGTGTTATTT 9088
Qy 782 GAGATAAGTCTTTTTCATGTTTTTCCCTATGAAATAAATTAATTTCCCTT 829
Db 9089 TTTTGTAGTTTTTTTTTTTTTTTTTTTTTAAATTTTAGATTTTGTGTTATTTATTT 9136
```

Search completed: August 15, 2004, 01:59:51
Job time : 495.203 secs

FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204.708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 89
LENGTH: 5844
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-89

Query Match 5.7%; Score 49.2; DB 4; Length 5535;
Best Local Similarity 27.4%; Pred. No. 0.013;
Matches 153; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 276 TTCTAAAGTGTATTAACTATTATATGTCAAATTTGAAACAGCTCTTCACTAC 335
Db 1262 TTATATTTTCTGTTATTTTAAATGGGTGTATGTTTATTATTATTAAATTT 1321
QY 336 TAGCAATTTATATCAGCACTTGTATTGTCACACTTTTAAAGTTTTCATTCATG 395
Db 1322 TAGGGTATTTTATGATTTTGAATTTTAAATTTTATTATTTATTTTAAATTT 1381
QY 396 NNN 455
Db 1382 TTATGTTTATTTGGGTTTAAATTTAGTTATTTTGGTATAGAAGTTATTGTTT 1441
QY 456 NNN 515
Db 1442 ATCGGTAATTTATTAATTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 1501
QY 516 NNN 575
Db 1502 TTGTTATATTTAGGTTATATGATAAATTTTCTGTTTCTGTTTCTGTTTCTG 1561
QY 576 NNN 635
Db 1562 GACGGAGTTTCGTTTATTTATTTAGTTGAGTGTAAATGCGCGATTTTAGTTT 1621
QY 636 ATTTAGTTTCTCATTTAAACCTGTTTAAATATTTTCAAAATCGACATTATGAT 695
Db 1622 ATTTTCTTCTCGGTTTAAAGCATTTTCTGTTTCTGTTTCTGTTTCTGTTTCT 1681
QY 696 ATATTATCTTCCAAATTTTAAATTTTCTGTTTCTGTTTCTGTTTCTGTTTCT 755
Db 1682 TGCTTATTTATGTTTCGTTTAAATTTTCTGTTTCTGTTTCTGTTTCTGTTTCT 1741
QY 756 CTGGAATCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 815
Db 1742 ACGGGGTTTATTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1801
QY 816 TAAATTTTCTTCTCTGTA 834
Db 1802 GTTATTTTGATTTTATA 1820

RESULT 4
US-10-204-708-89
Sequence 89, Application US/10204708
Patent No. 667731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation

FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204.708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 89
LENGTH: 5844
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-89

Query Match 5.6%; Score 48.6; DB 4; Length 5844;
Best Local Similarity 27.6%; Pred. No. 0.018;
Matches 159; Conservative 0; Mismatches 417; Indels 0; Gaps 0;

QY 275 TTCTAAAGTGTATTAACTATTATATGTCAAATTTGAAACAGCTCTTCACTTA 334
Db 4031 TATTGATATAGTTATATGTTAGTTATTTTATTGTTTGAATATATTTATGATCG 4090
QY 335 CTAGCAATTTATATCAGCAACACTTGTATTGTCAGACTTTTAAAGTTTTCATTCAC 394
Db 4091 GAAAGAAAGTTTATGATTTTAAAGTAGTTAGTTTATTTTATTTTATTTATTTA 4150
QY 395 NNN 454
Db 4151 TTTATATATTTTATGAGGTTTCGTTTCTGTTGTTAGTTAGTTAGTTAGTTAG 4210
QY 455 NNN 514
Db 4211 TCGGTTTATGTAATTTTCGTTTTCGGGTTAAAGTTATTTTGTGTTTATTTGA 4270
QY 515 NNN 574
Db 4271 GTAGTGGGATTATAGCGGTAGGTCGTTACGTTCCGTTAAATTTTCTGTTTCT 4330
QY 575 NNN 634
Db 4331 TTAGTAGAGATGGGGTTTATATGTTCTGTTTAGGTTGCTTTCGAATTTTTCG 4390
QY 635 TATTATGTTTCTCATTTTAAACCTGTTTAAATATTTTCAAAATCGACATTATGA 694
Db 4391 AGTTTATTTTATTTAGTTTATTTTAAAGTGTAGGATTATAGGTTATCGTTAT 4450
QY 695 TATATATCTCCAAATTTTAAATTTTCTGTTTCTGTTTCTGTTTCTGTTTCTG 754
Db 4451 TATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4510
QY 755 GCTGGAATTCATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 814
Db 4511 TTAGGTTGGAGTGTAGTCACGAGATTTTAGTTTATTTAGTTTATTTTATTT 4570
QY 815 ATAAATTTATTTCTCTCTGTTATTCGAGGTAGCTGGA 850
Db 4571 GCGATTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4606

RESULT 5
US-09-426-290-1/c
Sequence 1, Application US/09426290
Patent No. 6410712
GENERAL INFORMATION:
APPLICANT: Berglind Ran Olafsdottir

APPLICANT: Jeffrey Gulcher
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345-2001-000
CURRENT APPLICATION NUMBER: US/09/426,290
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match 5.6%; Score 48.2; DB 4; Length 168575;
Best Local Similarity 59.28; Pred. No. 0.044;
Matches 100; Conservative 0; Mismatches 68; Indels 1; Gaps 1;
Qy 11 TCATTCACTTAAAGTTTATGAGATATATCCATGTTGAATTTGTAGCTGTGTTGATG 70
Db 13525 TCACTTAGCTTAATATTTTCAAGGTTCTTCCATGTTTATGTCGTATCCGTACTTTATTA 13466
Qy 71 ATTTTACATATATAGTATATACATCCATGAGTAGTTCAGTAGATA-ATCCTCCTAT 129
Db 13465 TTTTGTGGCTGAATAGTATTTATATATAGATATGCCACATTTATCTATTCAGCAT 13406
Qy 130 TGTTAACATTTGGTTGCTTCATTTGACCTATTTTAAACAGGCCT 178
Db 13405 TGATGAATTTGGATTTGTTTCCATTTTGGCAATATATATAAATACT 13357

RESULT 6
US-08-559-896B-1/c
Sequence 1, Application US/08559896B
Patent No. 6310046
GENERAL INFORMATION:
APPLICANT: Patrick E. Duffy
APPLICANT: Christian F. Ockenhouse
TITLE OF INVENTION: SEQUESTIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran
STREET: USA MRC - MCMR-JA
CITY: FORT DETRICK, FREDERICK
STATE: MARYLAND
COUNTRY: USA
ZIP: 21702-5012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.5
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,896B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Moran, John
REGISTRATION NUMBER: 26,313
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 619-2065
TELEFAX: (301) 619-7714
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1956 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
US-08-559-896B-1
Query Match 5.6%; Score 48; DB 4; Length 1956;
Best Local Similarity 26.5%; Pred. No. 0.02;
Matches 138; Conservative 0; Mismatches 383; Indels 0; Gaps 0;
Qy 305 TATGTCAAATTTGAACAGCTCTTTCACTTACTAGCAATTTATATACGACACTTGTTA 364
Db 1360 TGTCTAAATTTGTCACAACTTAATCAACTTCTACATTTTCATTATAATTTATTTGTG 1301
Qy 365 TTGTCAGACATTTTAAGTTTTCATTCCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 424
Db 1300 GCTTATTATTAAATTAATTCATCTCTTATTATTATTATTATTATTATTATTATCAT 1241
Qy 425 NNN 484
Db 1240 TATTGTCGTGATTATAAATACATTTTACATTTCTTCTTTATTTTGGTTAGATTTCAT 1181
Qy 485 NNN 544
Db 1180 TTTTAAATTTCTATGATAAATATTATCTTTTTCATTTCTGCTTAAATTTCTCATGGGA 1121
Qy 545 NNN 604
Db 1120 TTTTCATACCTGCTCTCTATCCATTTCTATCTAATCTCTTTTATAAGCTATCGATTT 1061
Qy 605 NNN 664
Db 1060 CTTCTAGTCTTAAATTTTGTATTTTCTTTTGTATATCTTGTAAATTTGCTCTCTCTTA 1001
Qy 665 GGTAAATATTTCAAAATCGACATTTATATATATCTTCCAAAATTTTATAATAATTTT 724
Db 1000 CATGTTGAATTTTCATCATCATCAATTTGTTTAAATTTCTTCTATATTTTATCTATT 941
Qy 725 GTCTTTTTCACATTTTAGTCTTTTAGCTCAGCTGGAAATTCATTTCTGTGTGTGTGTGAG 784
Db 940 GATCTTTATCCATTTTATTATTATTTCTCTTTTCAATTTTGTGTAATTTTCTCTATGAA 881
Qy 785 ATAAGTCTTTTTCATGTTTTCCTTATGAATAAATTTT 825
Db 880 TTTTCTTTTCTCTCTCTCTATTTTCTTAAATTTCTCT 840

RESULT 7
US-09-351-794A-1/c
Sequence 1, Application US/09351794A
Patent No. 6641815
GENERAL INFORMATION:
APPLICANT: DUFFY, PATRICK E.
APPLICANT: OCKENHOUSE, CHRISTIAN F.
TITLE OF INVENTION: SEQUESTIN
FILE REFERENCE: 38644-175519
CURRENT APPLICATION NUMBER: US/09/351,794A
CURRENT FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 08/559,896
PRIOR FILING DATE: 1995-11-17
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1956

```

; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-351-794A-1
Query Match 5.6%; Score 48; DB 4; Length 1956;
Best Local Similarity 26.5%; Pred. No. 0.02;
Matches 138; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

QY 305 TATGTCATTTTCAACAGCTCTTTCACCTACTAGCAATTAATATACAGCAACACTTGTTA 364
DB 1360 TGTCTAAATTCGTACAACTAATCAACTCTACATTTTCATTAATATTATTATTGTTG 1301
QY 365 TTCTCAGACTTTTAAAGTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 424
DB 1300 GCTTATTAATTAATTCACATCTATTAATTAATTAATTAATTAATTAATTAATTA 1241
QY 425 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 484
DB 1240 TATTGTCGTATTATAAATACATTTTACATTTTCTTCTTATTTTGGTTAGATTCAT 1181
QY 485 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 544
DB 1180 TTTTAATTTCTATGTAATAATATATCTTTTTCATTCGCTTAATTTCTCATGGGA 1121
QY 545 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 604
DB 1120 TTTTCATACCTTGTCTCTATCCATCTATCTAATTTCTTTTATAAAGTCTATCGATT 1061
QY 605 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 664
DB 1060 CTCTAGTCTAAATTTTGTATTTTCTTTTGTATTAICTGTAATATGCTCTCTTA 1001
QY 665 GGTAAATATTCACAAATCGACATTAATATATATCTTCCAAAATTTTAAATAATTT 724
DB 1000 CATGTTGAATTCATCACTATCCATTTTGTAAATTTCTTCTCATATATTTATCTATT 941
QY 725 GTCTTTTTCACATTTTAGTCTTTAGCTCAGCTGGAATTCATTTCTGCTGCTGTGAG 784
DB 940 GATCTTTTACATTTTATTTATTTTCTCTTTTCAATTTTGTGTAATTTTCTCTAGAA 881
QY 785 ATAGTCTTTTTCATGTTTTTCCCTATGAATAAATATTTT 825
DB 880 TTTTTCCTTTTCTCTCTGCTATTTTCTAATTTCTTCT 840

RESULT 8
US-09-339-333D-1/c
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.04TAUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452

; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
```


NAME/KEY: misc feature
LOCATION: 199122..201122
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 201123..201234
OTHER INFORMATION: exon S
NAME/KEY: exon
LOCATION: 201123..201560
OTHER INFORMATION: exon S2
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U
NAME/KEY: exon
LOCATION: 216836..216994
OTHER INFORMATION: exon V
NAME/KEY: exon
LOCATION: 216836..217077
OTHER INFORMATION: exon V2
NAME/KEY: exon
LOCATION: 217671..217764
OTHER INFORMATION: exon V1
NAME/KEY: exon
LOCATION: 227655..227736
OTHER INFORMATION: exon V4
NAME/KEY: exon
LOCATION: 238715..238919
OTHER INFORMATION: exon V3
NAME/KEY: exon
LOCATION: 240440..240673
OTHER INFORMATION: exon W
NAME/KEY: exon
LOCATION: 240440..241153
OTHER INFORMATION: exon W2
NAME/KEY: exon
LOCATION: 241072..241291
OTHER INFORMATION: exon X
NAME/KEY: exon
LOCATION: 244353..244561
OTHER INFORMATION: exon Y
NAME/KEY: exon
LOCATION: 246273..247802
OTHER INFORMATION: exon Z
NAME/KEY: misc feature
LOCATION: 247803..249803
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 8316
OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 21672
OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 65485
OTHER INFORMATION: 8-128-33 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 95396
OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 107281
OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 160640
OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 160876
OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 168974
OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 170810
OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 173358
OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 189357
OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 197163
OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 200778
OTHER INFORMATION: 8-303-235 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202851
OTHER INFORMATION: 8-300-221 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202679
OTHER INFORMATION: 8-300-193 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 203378
OTHER INFORMATION: 8-299-128 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204138
OTHER INFORMATION: 8-296-213 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204605
OTHER INFORMATION: 8-252-190 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 204934
OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 205206
OTHER INFORMATION: 8-295-248 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 205329
OTHER INFORMATION: 8-295-125 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 206064
OTHER INFORMATION: 8-293-130 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 206545
OTHER INFORMATION: 8-292-198 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 207313
OTHER INFORMATION: 8-251-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208285
OTHER INFORMATION: 8-289-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208960
OTHER INFORMATION: 8-287-249 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 209123
OTHER INFORMATION: 8-287-86 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 209631
OTHER INFORMATION: 8-285-319 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210361
OTHER INFORMATION: 8-283-278 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210463
OTHER INFORMATION: 8-283-176 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210486
OTHER INFORMATION: 8-283-153 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210583
OTHER INFORMATION: 8-283-56 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 210879

```

; OTHER INFORMATION: 8-282-345 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210964
; OTHER INFORMATION: 8-282-260 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 210979
; OTHER INFORMATION: 8-282-245 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 211050
; OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGA
; NAME/KEY: allele
; LOCATION: 211132
; OTHER INFORMATION: 8-282-92 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 211247
; OTHER INFORMATION: 8-281-367 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211315
; OTHER INFORMATION: 8-281-299 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211366
; OTHER INFORMATION: 8-281-248 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 212520
; OTHER INFORMATION: 8-279-197 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 212821
; OTHER INFORMATION: 8-278-289 : polymorphic base C or T
; NAME/KEY: allele

```

```

Query Match          5.4%; Score 46.6; DB 4; Length 319608;
Best Local Similarity 65.8%; Pred. No. 0.12;
Matches 98; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

QY 27 TTAAGATATATCCATGTTGAATTTTGTAGCTGGTTGATGATTTTAC-ATTATAT 85
Db 15622 TTTTGGATTTTCATCCATGTTGCTGGGTAGCAATAATCTACTCATTTTAACTGTGGTCT 15563

QY 86 AGTATTACATCCATGATAGTTCTCAGTAGAT-AATCTCTCTATTGTTTACATTTGG 144
Db 15562 AGAATCCATCTATGATATACCAAAATCCATCCATTTCTAGTTGATGACATTTGG 15503

QY 145 TTCTCTCTATTTTGACCTATTTTAAACA 173
Db 15502 CTTTTCCATTTGGAGCTATTTCCAAATA 15474

```

```

RESULT 10
US-08-951-871-1/c
; Sequence 1, Application US/08951871
; Patent No. 5866398
; GENERAL INFORMATION:
; APPLICANT: XU, SHUANG-YONG
; APPLICANT: XIAO, JIAN-PING
; TITLE OF INVENTION: METHOD FOR CLONING
; TITLE OF INVENTION: AND PRODUCING THE BELL RESTRICTION ENDONUCLEASE IN E.
; TITLE OF INVENTION: COLI
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,871
; FILING DATE:

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 978-927-5054
; TELEFAX: 978-927-1705
; TELEX:

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2739 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..2736
; OTHER INFORMATION:
; US-08-951-871-1

```

```

Query Match          5.4%; Score 46.2; DB 2; Length 2739;
Best Local Similarity 47.2%; Pred. No. 0.054;
Matches 141; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 91 TACATTCATGGATAGTTCCTCAGTAGATAATCCCTCTATTGTTTAAACATTTGGTTCCTT 150
Db 917 TAAATTTGTTTATGGAGAAAGTACATAATCCAAAGAAATTTATTTTCATCTTGGAAATCTA 858

QY 151 CTCATTTTGACCTATTTTAAACAGGCCCTCTTTAAATATACACTTCTCTGATGTATGCT 210
Db 857 TTAATAAATCTCTGATCTATATGAAATAAACAAGTCATACCTCTCTGTAGTAAGAGT 798

QY 211 AGAATGGAGTGGCTGAATAAAGTGGCTGAATCATCTTCAACTCTAGTAAGATGTCAA 270
Db 797 AGAATATCATTTCCCGGTCAAAATAAAGGATATTCATTAATATATTTATCAACATAC 738

QY 271 ACTGTTTCTAAAGTGTGTTTATTTTAACTATTATATGTCAATTTTGAACAGCTCTTTCA 330
Db 737 AATTTGATTCATGCTCTCAAGTTTAAAGCTATAGCATCTTGTGATATTTCTTTTC 678

QY 331 CTTACTAGCAATTATATATCAGCAACACTTGTATTGTGACACTTTTAAAGTTTTCATTC 389
Db 677 ATTAATTCATATACACTTTTCATTAAACATCTTGATGGCTAAACAGCAGTAATGACTC 619

```

```

RESULT 11
US-10-204-708-6
; Sequence 6, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01

```



```

; Patent No. 6593114
;
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;           Gil H. Choi
;           Patrick S. Dillon
;           Craig A. Rosen
;           Steven C. Barash
;           Michael R. Fannon
;
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;
; NUMBER OF SEQUENCES: 5256
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER-READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hymen
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
;
; INFORMATION FOR SEQ ID NO: 521:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 521:
US-08-956-171E-521

```

[illegible]

D	b	332	ATTCTAAGTCGCACCTTTAAATCGTCTCCTCTCATTTAAAGTCCTTTCAATTCAAGGCTTAATCTT
Q	y	518	NN
D	b	272	TTTTTATATCTTCAATTCCTTTTAAGAACTCCTCACCTTGTTTTCAAAAGTATTAATA
Q	y	578	NN
D	b	212	CACCTTTCATATCTTTTCATGAGATAATCCTAGATAATCTAAATATCAGTATTTCCATTTT
Q	y	638	TATATGTTTTGTCATTTAAAAAACCTGGTTAAATATTTACAAAATCGACATTATGATAT
D	b	152	TTATTTTTTCTCTTCTTCACTAAAAAATTTCTTCAAAGCCTCATCTKAGCCATAAATCTCT
Q	y	698	ATT 700
D	b	92	GTT 90
RESULT 15			
US-10-204-708-20			
; Sequence 20, Application US/10204708			
; Patent No. 667731			
; GENERAL INFORMATION:			
; APPLICANT: OLEK, Alexander			
; APPLICANT: FIENBROCK, Christian			
; APPLICANT: BERLIN, Kurt			
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication			
; TITLE OF INVENTION: by Assessing DNA Methylation			
; FILE REFERENCE: 5013.1012			
; CURRENT APPLICATION NUMBER: US/10/204,708			
; CURRENT FILING DATE: 2003-05-06			
; PRIOR APPLICATION NUMBER: PCT/EP01/03971			
; PRIOR FILING DATE: 2001-04-06			
; PRIOR APPLICATION NUMBER: DE 10019058.8			
; PRIOR FILING DATE: 2000-04-06			
; PRIOR APPLICATION NUMBER: DE 10019173.8			
; PRIOR FILING DATE: 2000-04-07			
; PRIOR APPLICATION NUMBER: DE 10032529.7			
; PRIOR FILING DATE: 2000-06-30			
; PRIOR APPLICATION NUMBER: DE 10043826.1			
; PRIOR FILING DATE: 2000-09-01			
; NUMBER OF SEQ ID NOS: 98			
; SEQ ID NO 20			
; LENGTH: 6865			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)			
US-10-204-708-20			

Query Match	5.3%	Score 45.6	DB 4	Length 6866
Best Local Similarity	31.9%	Pred. No. 0.09		
Matches 258	Conservative 0	Mismatches 547	Indels 4	Gaps 2
QY	29	ATGAGATATATCCATGTTGAATTTTGTAGCTGTGGTTTG-ATGATTTTACATTATATAG	87	
Db	4867	ATTAGATTTTGTATTTTAGTATATCTTTTTTGTGTTGTGATTTTTTTATTATGAAAT	4926	
QY	88	TATTACATTCATGGATAGTCTCAGTAGATAATCCTCTATTGTTTAAACATTTGCGTGG	147	
Db	4927	TTATATTGTTTATATAATAAATGAAGAAAAAATTTTGTATTGTATATAAATATT	4986	
QY	148	CTTCTCATTTTGGACCTATTTTAAACAGGCCTCTTTAAATATACACTCTCTGTAGTGTAT	207	
Db	4987	TATTAGTATATAAAAAATATATTTTGTGAATTTTAGATTTTAAATGTGTGTGGGTGATA	5046	
QY	208	GCTAGAAATGGAGTGGCTGGAATAAAGTGGCTGAATCATCTTCAACTCTAGTAAAGATGT	267	
Db	5047	TATTTAATGTTAATGTTGGTATTTGGGGAAGTAGATTTTATAAATGTGATAGGTTGT	5106	
QY	268	CAAACTCTTTTCTAAAAGTGTTTTATTTTAACTATTTATATGTCAAATTTTGAACAGCTCTT	327	

```
Db 5107 GGTATATATTTTATTTTGTGTTGGTAAAGTTTTTAATTGAGTGAAGATTTTT 5166
Qy 328 TCCTTACTAGCAATTTATTATCAGACACTTGTATTGTGACACTTTTAAAGTTTCAT 387
Db 5167 TTTTGTGTTTGTGTTTGTGAGTAGGTTTATTTT--GTCGTTTAGGTTGGAGT 5223
Qy 388 TCACTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 447
Db 5224 GTAGTGGTATCAATTTGTTTATTGTTAATTCGGTTTTTTGGGTTTAAATAATTCGTG 5283
Qy 448 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 507
Db 5284 TTTTAGTTTTTTGAGTAGTGGGATTATAGGTATATATATATGTTAGTTAATTTGT 5343
Qy 508 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 567
Db 5344 ATTATTAGTAGAGTTGGGTTTCGTTATGTTGTTAGTGGTTTGAATTTTGTATTT 5403
Qy 568 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 627
Db 5404 AAGTAATTTGTTTGTGTTTAGTTTTTAAAGTGTGAGATTATAGGTATGAGTTATT 5463
Qy 628 GTGATCATATTTATGTTTGTCTCATTTTAAAAAACCTGTTTAAATATTTTCAAAATCGAC 687
Db 5464 TTAGTTAAAGTTTGTGTTTATATTTTATGTTTGTGTTTGTGTTTGTGTTTAAAT 5523
Qy 688 ATTAGATATATATCTCCAAAATTTTAAATTTTGTCTTTTTTCCACATTTTAGTCTT 747
Db 5524 TATTTTAAATCAGGATTTTAAATTTATTTATTAGAAATTTAAATTTTATTATTTT 5583
Qy 748 TAGCTCAGCTGGAATTCATTTCTGTGTGTGTGTGAGATAAGTCTTTTCATGTTTCC 807
Db 5584 TTAATTTGGTTTTTGTGTTTGTATTTTGTGTTTATTTTATAGTTTTTATTGTTT 5643
Qy 808 CTATCAATAAATTTATTTCTCTCTGATT 836
Db 5644 TATAGAGTTTTTTGATTTTATGTTTATT 5672
```

Search completed: August 14, 2004, 20:48:09
Job time : 85.9109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 2442.69 Seconds
(without alignments)
9989.844 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563
Sequence: 1 ggtacatcttggtgtggtat.....gcgcgtagctctgcagttctag 563

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb.ba.*
2: gb.btg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vit.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	511.4	90.8	185566	9	AC004168	AC004168 Homo sapi
C 2	53.2	9.4	7218	6	166494	166494 Sequence 14
C 3	49.6	8.8	128683	2	AC135793	AC135793 Oryza sat
C 4	48.6	8.6	148526	10	AL732503	AL732503 Mouse DNA
C 5	48.2	8.6	94391	2	AC135575	AC135575 Rattus no
C 6	48.2	8.6	168088	2	AC110880	AC110880 Mus muscu
C 7	48.2	8.6	168773	10	AL731820	AL731820 Mouse DNA
C 8	47.8	8.5	137648	2	AC024316	AC024316 Homo sapi
C 9	47.8	8.5	158608	9	AC051635	AC051635 Homo sapi
C 10	47.8	8.5	193978	2	AC068678	AC068678 Homo sapi
C 11	47.8	8.5	209462	2	AC011814	AC011814 Homo sapi
C 12	47.6	8.5	207777	2	AC102621	AC102621 Mus muscu
C 13	47.6	8.5	222906	2	AC132603	AC132603 Mus muscu
C 14	47.4	8.4	161021	10	EX537253	EX537253 Mouse DNA
C 15	47	8.3	220394	2	AC133842	AC133842 Rattus no
C 16	46.8	8.3	214285	10	AC107635	AC107635 Mus muscu
C 17	46.8	8.3	222845	10	AC100491	AC100491 Mus muscu
C 18	46.6	8.3	65772	2	AC123762	AC123762 Mus muscu
C 19	46.6	8.3	120507	10	AL928895	AL928895 Mouse DNA
C 20	46.6	8.3	177955	2	AC134460	AC134460 Mus muscu
C 21	46.6	8.3	178455	2	AC137969	AC137969 Mus muscu
C 22	46.6	8.3	198829	10	AC120548	AC120548 Mus muscu
C 23	46.6	8.3	200487	2	AC145577	AC145577 Mus muscu
C 24	46.6	8.3	221362	2	AC134084	AC134084 Rattus no
C 25	46.6	8.3	236639	2	AC103213	AC103213 Rattus no
C 26	46.6	8.3	245711	2	AC098021	AC098021 Rattus no
C 27	46.6	8.3	253401	2	AC097127	AC097127 Rattus no
C 28	46.6	8.3	268574	2	AC087129	AC087129 Mus muscu
C 29	46.6	8.3	282426	2	AC145569	AC145569 Mus muscu
C 30	46.2	8.2	157322	9	AC012103	AC012103 Homo sapi
C 31	46.2	8.2	173324	2	AC016038	AC016038 Homo sapi
C 32	46.2	8.2	187557	2	AC069032	AC069032 Homo sapi
C 33	46	8.2	168580	10	AC127696	AC127696 Mus muscu
C 34	46	8.2	179455	2	AC102304	AC102304 Mus muscu
C 35	46	8.2	180017	2	AC115163	AC115163 Rattus no
C 36	46	8.2	191312	10	AL732420	AL732420 Mouse DNA
C 37	46	8.2	201923	10	AC121772	AC121772 Mus muscu
C 38	46	8.2	202010	2	AC136558	AC136558 Rattus no
C 39	46	8.2	202342	10	AL671867	AL671867 Mouse DNA
C 40	46	8.2	209438	2	AC142275	AC142275 Mus muscu
C 41	46	8.2	231704	10	AC051638	AC051638 Mus muscu
C 42	46	8.2	239439	10	AC117573	AC117573 Mus muscu
C 43	46	8.2	243425	2	AC108643	AC108643 Rattus no
C 44	46	8.2	255563	2	AC125834	AC125834 Rattus no
C 45	46	8.2	302176	2	AC096479	AC096479 Rattus no

ALIGNMENTS

RESULT 1
AC004168/c
LOCUS AC004168 Homo sapiens chromosome 4 clone B159L21 map 4q25, complete
DEFINITION sequence.
ACCESSION AC004168
VERSION AC004168.2 GI:4454440
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185566)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission

185556 bp DNA linear PRI 19-MAR-1999

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 185566)
AUTHORS      Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE        Direct Submission
JOURNAL      Submitted (23-FEB-1998) Department of Genetics, Stanford Human
              Genome Center, 955 California Avenue, Palo Alto, CA 94304, USA
REFERENCE    3 (bases 1 to 185566)
AUTHORS      Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
TITLE        Direct Submission
JOURNAL      Submitted (19-MAR-1999) Department of Genetics, Stanford Human
              Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
COMMENT      On Mar 19, 1999 this sequence version replaced gi:3603082.
              Quality: Phrap Quality >=40 100.0% of Sequence;
              Estimated Total Number of Errors is 0.0.
              STS Content:
              SHGC-50970 G33969
              SHGC-23829 G28472
              WI-30731 G23237
              SHGC-23845 G28477
              SHGC-23843 G28476
              SHGC-50971 G34439.
              Location/Qualifiers
                1. 185566
                  /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
                  /chromosome="4"
                  /map="4q25"
                  /clone="B159L21"
                  /clone_lib="Caltech/Research Genetics Whole-Genome BAC
                  Library"

FEATURES             source
    source            1. 185566
    Location/Qualifiers
      1. 185566
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4q25"
        /clone="B159L21"
        /clone_lib="Caltech/Research Genetics Whole-Genome BAC
        Library"

ORIGIN
Query Match      90.8%; Score 511.4; DB 9; Length 185566;
Best Local Similarity 93.6%; Pred. No. 6.3e-109;
Matches 523; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GGTACATCTTGGCTGTGGATGGAAATTTGACATACATCTTTTATTTTCTTTACCTGTTCAT 60
DB 25328 GGTACATCTTGGCTGTGGATGGAAATTTGACATACATCTTTTATTTTCTTTACCTGTTCAT 25269

QY 61 ATCAAACTTAGGATGATTTACTTCCAAAGCGGTTAACTTATTCAGAGTTGACTGAGTC 120
DB 25268 ATCAAACTTAGGATGATTTACTTCCAAAGCGGTTAACTTATTCAGAGTTGACTGAGTC 25209

QY 121 TCCTATTTTCTTAAATTTACTAGATGAGGCTCCAGAACTACAGAAATAGAGGAA 180
DB 25208 TCCTATTTTCTTAAATTTACTAGATGAGGCTCCAGAACTACAGAAATAGAGGAA 25149

QY 181 AGTCTCCATTGAGCCATGAACTGTGAGCACCTGGCATTAAAGCATGAAGATAGGCTTC 240
DB 25148 AGTCTCCATTGAGCCATGAACTGTGAGCACCTGGCATTAAAGCATGAAGATAGGCTTC 25089

QY 241 TATGGTAGGACTGGAGTAGGAGCATTCAGGAAAGGATCTCAGAGGTCAGAAACAATA 300
DB 25088 TATGGTAGGACTGGAGTAGGAGCATTCAGGAAAGGATCTCAGAGGTCAGAAACAATA 25029

QY 301 GATTATCAGTAAATACCTCTGGACCAAGAAAGACCTTGAATCTCGGTGATGTA 360
DB 25028 GATTATCAGTAAATACCTCTGGACCAAGAAAGACCTTGAATCTCGGTGATGTA 24969

QY 361 TACATATTTAATACAAATGAGAGCTCTTGGAGGTGAGTGAGAGTATAGGGCTGAAG 420
DB 24968 TACATATTTAATACAAATGAGAGCTCTTGGAGGTGAGTGAGAGTATAGGGCTGAAG 24909

QY 421 AACAGGAGACAGAAGCAATTCATTAATGTCTCTACACAGAAAGCTGATTATAGATAATG 480
DB 24908 AACAGGAGACAGAAGCAATTCATTAATGTCTCTACACAGAAAGCTGATTATAGATAATG 24849

QY 481 GTACATTCATATAAATGATGATATTTACTATTAATAAACAATACCA 525
DB 24848 GTACATTCATAT-AACATGATATATTTACTATTAATAAACAATAACA 24805

```

```

RESULT 2
LOCUS      I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION    I66494.1 GI:2724471
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 7218)
AUTHORS    Dorrer,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE      Recombinant Fowlpox virus
JOURNAL    Patent: US 5670367-A 14 23-SEP-1997;
FEATURES   Location/Qualifiers
            source
              1. 7218
                /organism="unknown"
                /mol_type="unassigned DNA"

ORIGIN
Query Match      9.4%; Score 53.2; DB 6; Length 7218;
Best Local Similarity 1.9%; Pred. No. 0.033;
Matches 7; Conservative 218; Mismatches 141; Indels 0; Gaps 0;

QY 135 AATTCTAGAGTGAGGCTCCAAGAACTACAGAAATAGAGAAAGTCTCCATTGAGC 194
DB 1444 AATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1385

QY 195 CATGAACGTGTCGACACCTGCGCATTTAAGCATGAGAGTAGGCTTCTATGTTAGGACTG 254
DB 1384 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1325

QY 255 GAGTAGGAGCATTCAGAGAAAGATCTCAGAGGTCAGAAACAATAGATTATTCAGTTAAA 314
DB 1324 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1265

QY 315 TACTTCGGACCAAGAGAGACCTTGAATCTCGCTCGGTGATGATATATATATATAC 374
DB 1264 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1205

QY 375 ACAATCAAGAGCTCTTGGAGGTGAGTGAGGCTGATAGGCTCAAGAACAGAGACAGAA 434
DB 1204 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1145

QY 435 GACAACTCAATGCTCTTACAGAAAGCTGATTATATAGATATGATGATCATCATATAA 494
DB 1144 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1085

QY 495 ACATGA 500
DB 1084 RRRRR 1079

RESULT 3
LOCUS      AC135793
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 11 clone
ACCESSION  Oryza sativa (japonica cultivar-group) chromosome 11 clone
VERSION     Oryza sativa (japonica cultivar-group) chromosome 11 clone
KEYWORDS    HTG; HTGS PHASE2; HTGS ACTIVEFIN.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE    1 (bases 1 to 128683)
AUTHORS      Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
              Overton II,L., Tsitrin,T., Kim,M., Bera,J., Jin,S., Fadrosch,D.W.,
              Tallon,L., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S.,
              Riedmuller,S.B., Utterbach,T., Feldblyum,T., Yang,Q., Haas,B.,
              Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and

```



```

Fraser, C.
Oryza sativa ssp. japonica cv. Nipponbare TIGR CLONE BAC genomic
sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128683)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 128683)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* *
* 1 128683: contig of 128683 bp in length.
* Location/Qualifiers
* 1..128683
* /organism="Oryza sativa (japonica cultivar-group)"
* /mol_type="genomic DNA"
* /cultivar="Nipponbare"
* /db_xref="taxon:39947"
* /chromosome="11"
* /clone="OSUNB50017P15"
* /note="japonica cultivar-group"
*
FEATURES
source
1..128683
Query Match
Best Local Similarity 8.8%; Score 49.6; DB 2; Length 128683;
Matches 60; Conservative 232; Mismatches 208; Indels 4; Gaps 2;
QY 32 ATACTTTTATTTTCTTACCTGTATCATATCAATCTTAGAGTGTATTCTTCAAGGC 91
DB 50753 RMYRKTCTKWKYKSTTTGCAAMWTAGKMYGKSWTGTGSCVKKRYRWAYCYKWSMK 50694
QY 92 GGTAAACTTATTCAGATTACACTGAGTCTCTTATTTCTCTTAAATTTACTAGAGTGAG 151
DB 50693 RMYKSYATRYMYRSWYTSYKMYRSGWYKMYRMYRMYRMYRMYRMYRMYRMYRMYR 50634
QY 152 GCTCAAGAACTACAGAAATAGAGAAAGTCTCTTATTCAGGCAATGATGTGAGCAAC 211
DB 50633 MMWYCRMWMAAWMYRMTKRGTTGASWKMWSASAAKYAWYCRMKRRKWKASYTRMK 50574
QY 212 TGGATTTTAAGCATGAGAGTAGGGCTTCTATGTTAGGACTGGAGTAGCAGCATTTCA 271
DB 50573 WKRMWATWYRGRARWYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 50516
QY 272 GGAAGAGTCTCAGAGTTCAGAAACAATAGATATATCATGTTAAATATCTTCTGGACCA 331
DB 50515 -RMSASSRKKWMMWMSWRSRYWMSWMSWMSWMSWMSWMSWMSWMSWMSWMSWMS 50457
QY 332 AGACCTTGAATCTCGCTCGGTGATGATATATATATATATATATATATATATATATAT 391
DB 50456 GKMWAGTKAKSMYTRSAWMTYWSWKMWYWGTTATKSYWTKWAKRAWWTKTATY 50397
QY 392 GAGGTGAGTGTAGAGTGTATAGGGTGTGAAGAACAGAGACAGACAGAAATTCATCTCT 451
DB 50396 YCRYATTATWTTCTTAGSMW -WYTYRKMCMEMRTWAARKSYTARSYKTKRSKSK 50338
QY 452 TACACAGAAAGTGAATATATAGTATATATGTTACATTCATATAACATGATATATTACT 511
DB 50337 RYRMRGKYARWKSRYRRRKKWTTGYGWCRRYWSYWRMTAMSAWMTAAWMYRKA 50278
QY 512 ATTAACAATACCAACACACACA 535

```

```

Db 50277 MWCCARCAATAAKASYAKRWTWW 50254
RESULT 4
LOCUS AL732503
DEFINITION Mouse DNA sequence from clone RP23-2809 on chromosome 4, complete
sequence.
ACCESSION AL732503
VERSION AL732503.8 GI:21998252
KEYWORDS HTG
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Pearce, A.
Direct Submission
Submitted (25-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Jul 26, 2002 this sequence version replaced gi:21911625.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1..148526
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-2809"
/clone_lib="RPI-23"

```

FEATURES

source

ORIGIN

```

Query Match 8.6%; Score 48.6; DB 10; Length 148526;
Best Local Similarity 64.9%; Pred. No. 0.23;
Matches 72; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 427 AGACAGAAGCAATTCATATGCTCTTACACAGAAGACTGATTATATAGATATGTTACAT 486
DB 135971 AGCCAGAAAAAATCCAGATGCTCTTCAACAGAGGATGTGTACAGAAATATGTTACAT 136030
QY 487 TCATATAAACAATGATATATTTCTAATTAATAAACATCCACACACACACA 537
DB 136031 TTACACAATGGAGTATTACTTGTGCTATTAAAAAATGAATTCATGAACA 136081

```

RESULT 5
 AC135575
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC135575 94391 bp DNA linear HTG 19-OCT-2002
 Rattus norvegicus clone CH230-20316, *** SEQUENCING IN PROGRESS
 ***, 52 unordered pieces.
 AC135575
 HTG: HTGS PHASE1
 Rattus norvegicus (Norway rat)
 Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 94391)
 Murny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gbregregoris, E., Geer, K., Gill, R., Grady, N., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Loresueta, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, N., Mundaas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G.,
 Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pflankoch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiecek, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished

2 (bases 1 to 94391)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

1118: contig of 1118 bp in length
 1119 1218: gap of unknown length
 1219 2501: contig of 1283 bp in length
 2502 2601: gap of unknown length
 2602 3811: contig of 1216 bp in length
 3818 3917: gap of unknown length
 3918 5144: contig of 1227 bp in length
 5145 5244: gap of unknown length
 5245 6260: contig of 1016 bp in length
 6261 6360: gap of unknown length
 6361 7588: contig of 1225 bp in length
 7589 7688: gap of unknown length
 7689 8694: contig of 1009 bp in length
 8695 8794: gap of unknown length
 8795 10206: contig of 1412 bp in length
 10207 10306: gap of unknown length
 10307 11520: contig of 1214 bp in length
 11521 11620: gap of unknown length
 11621 13019: contig of 1399 bp in length
 13020 13119: gap of unknown length
 13120 14579: contig of 1460 bp in length
 14580 14679: gap of unknown length
 14680 16092: contig of 1413 bp in length
 16093 16192: gap of unknown length
 16193 17509: contig of 1317 bp in length
 17510 17609: gap of unknown length
 17610 18996: contig of 1387 bp in length
 18997 19096: gap of unknown length
 19097 20159: contig of 1063 bp in length
 20160 20259: gap of unknown length
 20260 21714: contig of 1455 bp in length
 21715 21844: gap of unknown length
 21845 23289: contig of 1475 bp in length
 23290 23389: gap of unknown length
 23390 25099: contig of 1710 bp in length
 25100 25199: gap of unknown length
 25200 26334: contig of 1135 bp in length
 26335 26434: gap of unknown length
 26435 27867: contig of 1433 bp in length
 27868 27967: gap of unknown length
 27968 29046: contig of 1079 bp in length
 29047 29146: gap of unknown length
 29147 30488: contig of 1342 bp in length
 30489 30589: gap of unknown length
 30590 32259: contig of 1671 bp in length
 32260 32359: gap of unknown length
 32360 33897: contig of 1538 bp in length
 33898 35361: contig of 1364 bp in length
 35362 35361: contig of 1364 bp in length

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 52 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

* 35362 35461: gap of unknown length
* 35462 contig of 1453 bp in length
* 36914: gap of unknown length
* 36915 contig of 1649 bp in length
* 37015 38663: gap of unknown length
* 38664 39847: contig of 1084 bp in length
* 39848 39947: gap of unknown length
* 39948 41838: contig of 1891 bp in length
* 41839 41938: gap of unknown length
* 41939 44158: contig of 2220 bp in length
* 44159 44258: gap of unknown length
* 44259 45596: contig of 1328 bp in length
* 45597 45686: gap of unknown length
* 45687 47721: contig of 2035 bp in length
* 47722 47821: gap of unknown length
* 47822 49216: contig of 1395 bp in length
* 49217 49316: gap of unknown length
* 49317 51012: contig of 1696 bp in length
* 51013 51112: gap of unknown length
* 51113 53593: contig of 2481 bp in length
* 53594 53693: gap of unknown length
* 53694 55129: contig of 1436 bp in length
* 55130 55229: gap of unknown length
* 55230 56905: contig of 1676 bp in length
* 56906 57005: gap of unknown length
* 57006 58201: contig of 1196 bp in length
* 58202 58301: gap of unknown length
* 58302 60258: contig of 1957 bp in length
* 60259 60358: gap of unknown length
* 60359 61875: contig of 1517 bp in length
* 61876 61975: gap of unknown length
* 61976 63120: contig of 1145 bp in length
* 63121 63220: gap of unknown length
* 63221 65263: contig of 2049 bp in length
* 65264 65363: gap of unknown length
* 65364 67808: contig of 2439 bp in length
* 67809 67908: gap of unknown length
* 67909 71132: contig of 3224 bp in length
* 71133 71233: gap of unknown length
* 71234 72404: contig of 2172 bp in length
* 72405 73504: gap of unknown length
* 73505 76370: contig of 2866 bp in length
* 76371 76470: gap of unknown length
* 76471 78273: contig of 1803 bp in length
* 78274 78373: gap of unknown length
* 78374 82072: contig of 3699 bp in length
* 82073 82172: gap of unknown length
* 82173 84748: contig of 2575 bp in length
* 84749 84848: gap of unknown length
* 84849 87309: contig of 2462 bp in length
* 87310 87409: gap of unknown length
* 87410 89644: contig of 2235 bp in length
* 89645 89744: gap of unknown length
* 89745 94391: contig of 4647 bp in length.
FEATURES
    Location/Qualifiers
        source          1..94391
                        /organism="Rattus norvegicus"

Query Match      8.6%; Score 48.2; DB 2; Length 94391;
Best Local Similarity 61.6%; Pred No. 0.31;
Matches 77; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACAAATTCAAATGCTTACACAGAACTGATTATAGATA 477
DB 31033 AATGACCAAGAGCTGGAAAGAAATGTTGATGTCCTTCAACAGAGGAAAGAAATATAGAAAAC 31092
QY 478 ATGGTACATTCATATACATCATATATTTACTATTAATTAACACATACCAACACACACACA 537
DB 31093 GTGGTACATTCATATACATCATATTTACTATTAATTAACACATACCAACACACACACA 537
QY 538 CCTCG 542
DB 31153 CCTAG 31157

```

RESULT 6 AC110880 LOCUS

DEFINITION

ACCESSION VERSION KEYWORDS SOURCE

ORGANISM

REFERENCE AUTHORS TITLE JOURNAL REFERENCE

AUTHORS

AC110880 168088 bp DNA linear HTG 22-OCT-2002
Mus musculus clone RP24-456G3, WORKING DRAFT SEQUENCE, 3 ordered
pieces
AC110880
AC110880.4 GI:24211407
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 168088)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP24-456G3

2 (bases 1 to 168088)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,

Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,

Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,

Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,

Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,

McEwan,P., McKernan,K., Meidrim,J., Meneus,L., Mihova,T.,

Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Petta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 168088)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhtalter,B.,

Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,

Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafes,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,

Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,

Matthews,C., McCarthy,M., Meidrim,J., Meneus,L., Mihova,T.,

Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,

Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,

Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,

Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,

Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-OCT-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Oct 22, 2002 this sequence version replaced gi:20531822.

All repeats were identified using RepeatMasker:

Smitt,A.P.A. & Green,P. (1996-1997)

http://ftp.genome.washington.edu/RN/repeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L18747
 Center clone name: 456.G.3

----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 166916 bases at least Q40
 Consensus quality: 167532 bases at least Q30
 Consensus quality: 167776 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 167889; sum-of-contigs
 Quality coverage: 12.3 in Q20 bases; agarose-fp
 Quality coverage: 12.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

 * 1 43630: contig of 43630 bp in length
 * 43631 43730: gap of 100 bp
 * 43731 8102: contig of 39372 bp in length
 * 83103 83202: gap of 100 bp
 * 83203 168088: contig of 84886 bp in length.

FEATURES

source
 1. 168088
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-456G3"
 /clone_lib="RPCI-24 Male Mouse BAC"
 misc_feature
 1. 43630
 /note="assembly_fragment"
 misc_feature
 43731..83102
 /note="assembly_fragment"
 misc_feature
 83203..168088
 /note="assembly_fragment"

ORIGIN

Query Match 8.6%; Score 48.2; DB 2; Length 168088;
 Best Local Similarity 61.6%; Pred. No. 0.28; Indels 0; Gaps 0;
 Matches 77; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
 QY 418 AAGAACAGGACAGACAGCAATTCCTTACACAGAGACTGATTATATAGATA 477
 Db 162492 AATAGCCAGAGCTGGAAAGAACCCAGATGCTCCTCAACAGAGGAATGGATACAGAAAT 162551
 QY 478 ATGCTACATTCATATAACATGATATATTTACTAATAAACATACACACACACA 537
 Db 162552 ATGTGATTTATACATGTTGTACTACTCACCTATTAAACCAACAAATTTACAAAAT 162611
 QY 538 CCTCG 542
 Db 162612 CTTAG 162616

RESULT 7
 AL731820
 LOCUS AL731820 168773 bp DNA linear ROD 26-JUL-2002
 DEFINITION Mouse DNA sequence from clone RP23-135H19 on chromosome X, complete
 sequence.
 ACCESSION AL731820
 VERSION AL731820
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 168773)
 Griffiths, C.
 Direct Submission
 Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 29, 2002 this sequence version replaced gi:21955608.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW:
 SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-135H19 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.

FEATURES

source
 1. 168773
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-135H19"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 8.6%; Score 48.2; DB 10; Length 168773;
 Best Local Similarity 67.3%; Pred. No. 0.28;
 Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 418 AAGAACAGGACAGACAGCAATTCCTTACACAGAGACTGATTATATAGATA 477
 Db 82026 AATAGCCAGAACTGGAAAGAAATCCAGATGCTCCTTCAACAGAGATGGATACAGAAAT 82085
 QY 478 ATGCTACATTCATATAAACATGATATATTTACTAATAATTA 518
 Db 82086 GTGGTACATTTATATAAAGGAGTCTTATTCAGCTATTAAAA 82126

RESULT 8
 AC024316/c

LOCUS AC024316 137648 bp DNA linear HTG 26-SEP-2000
 DEFINITION Homo sapiens clone RP11-2608, WORKING DRAFT SEQUENCE, 17 unordered
 pieces.
 ACCESSION AC024316
 VERSION AC024316.3 GI:10305255
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 137648)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 137648)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepal,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenesor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Lagocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2000 this sequence version replaced gi:10198462.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4672
Center clone name: 26_O8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 128694 bases at least Q40
Consensus quality: 133079 bases at least Q30
Consensus quality: 135006 bases at least Q20
Insert size: 137000; agarose-fp
Insert size: 136048; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      3290: contig of 3290 bp in length
*      3291      3390: gap of 100 bp
*      3391      5203: contig of 1813 bp in length
*      5204      5304: gap of 100 bp
*      5304      7364: contig of 2061 bp in length
*      7365      7465: gap of 100 bp
*      7466      10939: contig of 3475 bp in length
*      10940      11039: gap of 100 bp
*      11040      15415: contig of 4376 bp in length
*      15416      15515: gap of 100 bp
*      15516      20495: contig of 4980 bp in length
*      20496      20595: gap of 100 bp
*      20596      25962: contig of 5367 bp in length

```

```

*      25963      26062: gap of 100 bp
*      26063      30774: contig of 4712 bp in length
*      30775      30874: gap of 100 bp
*      30875      36005: contig of 5131 bp in length
*      36006      36105: gap of 100 bp
*      36106      40573: contig of 4468 bp in length
*      40574      40673: gap of 100 bp
*      40674      47316: contig of 6643 bp in length
*      47317      47416: gap of 100 bp
*      47417      57551: contig of 10135 bp in length
*      57552      57651: gap of 100 bp
*      57652      67679: contig of 10028 bp in length
*      67680      67779: gap of 100 bp
*      67780      78221: contig of 10342 bp in length
*      78222      78221: gap of 100 bp
*      78222      105075: contig of 26854 bp in length
*      105076      105175: gap of 100 bp
*      105176      124687: contig of 19512 bp in length
*      124688      124788: gap of 100 bp
*      124789      137648: contig of 12861 bp in length.

```

FEATURES
source

```

1..137648
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="RP11-2608"
/clone="RP11-2608"

```

```

1..3290
/note="assembly_fragment"
vector_side:left

```

```

3391..5203
/note="assembly_fragment"

```

```

5304..7364
/note="assembly_fragment"

```

```

7465..10939
/note="assembly_fragment"

```

```

11040..15415
/note="assembly_fragment"

```

```

15516..20495
/note="assembly_fragment"

```

```

20596..25962
/note="assembly_fragment"

```

```

26063..30774
/note="assembly_fragment"

```

```

30875..36005
/note="assembly_fragment"

```

```

36106..40573
/note="assembly_fragment"

```

```

40674..47316
/note="assembly_fragment"

```

```

47417..57551
/note="assembly_fragment"

```

```

57652..67679
/note="assembly_fragment"

```

```

67780..78121
/note="assembly_fragment"

```

```

78222..105075
/note="assembly_fragment"

```

```

105176..124687
/note="assembly_fragment"

```

```

124788..137648
/note="assembly_fragment"

```

```

clone_end:17
vector_side:right

```

ORIGIN

```

Query Match      8.5%; Score 47.8; DB 2; Length 137648;
Best Local Similarity 61.8%; Pred. No. 0.36;
Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY      418 AAGACAGGAGACAGACAGACATTCAATGCTTACACAGAGACTGATTATATAGATA 477
|||||

```

Db 116769 AACGCCAGAAAAATGGAAACAATAACAAATGTCATCAACACAGAATGAATAAATAGTT 116771

QY 478 ATGTGTACATTTCATATAAACCATGATGATATATTTACTAAATTAACACATACCAACACACACAACA 537
|||||
Db 116709 GTGGTATATTTACACACAGAAAAAATGGCAGTAGAAGAACCTTACATTACACACACATCA 116650

QY 538 CCT 540
|||

Db 116649 CCT 116647

RESULT 9
LOCUS AC051635/c
DEFINITION Homo sapiens chromosome 18, clone RP11-540A4, complete sequence.
ACCESSION AC051635
VERSION AC051635.7 GI:17298634
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 158608)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158608)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,I., Boukhgalter,B., Brown,A., Burkett,G., Campioniano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,D., Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., MCPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stoljanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

DIRECT SUBMISSION
Submitted (15-Apr-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 158608)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,I., Boukhgalter,B., Brown,A., Camarata,J., Campioniano,A., Chang,J., Chazarot,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., McClean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,B., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., ROY,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stoljanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

```

/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-540A4"
/clone_lib="RP11-11 Human Male BAC"
8..106
/rpt_family="AluSg/x"
15..27
/note="<30 qual SNGL region"
75..77
/note="<30 qual SNGL region"
106..109
/note="<30 qual SNGL region"
112..228
/rpt_family="L1MB2"
complement(563..632)
/rpt_family="MIR"
1053..1057
/note="<30 qual SNGL region"
complement(1123..1352)
/rpt_family="MIR"
1624..1702
/rpt_family="L2"
complement(1853..2218)
/rpt_family="L2"
3091..3358
/rpt_family="L1MCS"
3357..3850
/rpt_family="L1MC1"
complement(3868..4006)
/rpt_family="LTR7"
4011..4167
/rpt_family="MIR3"
complement(4235..4560)
/rpt_family="LTR16C"
5345..5372
/rpt_family="TTTG)n"
5957..5984
/rpt_family="TTTG)n"
6132..6246
/rpt_family="MIR3"
6598..6718
/rpt_family="L1MB2"
6731..6769
/rpt_family="CA)n"
6940..7148
/rpt_family="L1MB2"
7153..7486
/rpt_family="L1MB8"
7510..8250
/rpt_family="L1MB2"
8279..8516
/rpt_family="L1MB2"
8612..8827
/rpt_family="AluJb"
9027..9126
/rpt_family="L1MB2"
9127..9435
/rpt_family="AluY"
9436..9486
/rpt_family="L1MB2"
9487..9805
/rpt_family="AluSx"
9806..9922
/rpt_family="L1MB2"
complement(9940..9989)
/rpt_family="MLT11"

Query Match      8.5%; Score 47.8; DB 9; Length 158608;
Best Local Similarity 61.8%; Pred. No. 0.35;
Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

418 AAGAACAGGACAGACAGACAAATGTCCTTACACAGAGACGTATTATAGATA 477

```

Db 134152 AACAGCCAGAAAATGGAACAATACAAATGTCCTTACACAGAAATGGATAAAGTT 134093
 QY 478 ATGGTACATTTCATATAACATCATATATTTACTTAATTTAAACATACACACACACA 537
 Db 134092 GTGGTATATTTACACACAGAAAAAATGTCAGTAGAAGACTTACATTCACACATCA 134033
 QY 538 OCT 540
 Db 134032 CCT 134030

RESULT 10
 AC068678
 LOCUS
 DEFINITION Homo sapiens chromosome 18 clone RP11-104N14 map 18, WORKING DRAFT
 SEQUENCE, 5 unordered pieces.
 AC068678 3 GI:16506981
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 1 (bases 1 to 193978)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 18, clone RP11-104N14
 2 (bases 1 to 193978)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellan,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 28, 2001 this sequence version replaced gi:13249473.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIER
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1295
 Center clone name: 104.N.14
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 4% of reads
 Sequencing vector: Plasmid; n/a; 96% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 191018 bases at least Q40
 Consensus quality: 192308 bases at least Q30
 Consensus quality: 192875 bases at least Q20

insert size: 190000; agarose-fp
 Insert size: 193578; sum-of-contigs
 Quality coverage: 7.3 in Q20 bases; agarose-fp
 Quality coverage: 7.1 in Q20 ba.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 59978: contig of 59978 bp in length
 * 59979 60078: gap of 100 bp
 * 60079 83937: contig of 23859 bp in length
 * 83938 84037: gap of 100 bp
 * 84038 113621: contig of 29584 bp in length
 * 113622 113721: gap of 100 bp
 * 113722 147302: contig of 33581 bp in length
 * 147303 147402: gap of 100 bp
 * 147403 193978: contig of 46576 bp in length.

FEATURES

source

Location/Qualifiers
 1..193978
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-104N14"
 /clone_lib="RPC1-11 Human Male BAC"
 misc_feature
 1..59978
 /note="assembly_fragment
 vector_side:left"
 clone_end:SP6
 misc_feature
 60079..83937
 /note="assembly_fragment"
 84038..113621
 /note="assembly_fragment"
 113722..147302
 /note="assembly_fragment"
 147403..193978
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 8.5%; Score 47.8; DB 2; Length 193978;
 Best Local Similarity 61.8%; Pred.No.0.34;
 Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 418 AAGACAGGAGACAGACAAATTCCTTACACAGAGACTGATTATATAGATA 477
 DB 101530 AACGCCAGAAATGGAAACATACAAATGTCCATCAACAGAAATGGATAAATAGTT 101589
 QY 478 ATGTACATTCATATAACATGATATATTTACTTAATTAACATACACACACACACA 537
 DB 101590 GTGGTATATTTACACACAGAAAAAATGGCAGTAGAAGAACTTACATTCACACATCA 101649
 QY 538 CCT 540
 DB 101650 CCT 101652

RESULT 11
 AC011814
 LOCUS AC011814 209462 bp DNA linear HTG 26-MAY-2000
 DEFINITION Homo sapiens chromosome 18 clone RP11-153D16 map 18, WORKING DRAFT
 SEQUENCE, 40 unordered pieces.
 ACCESSION AC011814
 VERSION AC011814.3 GI:8072439
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 209462)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 18, clone RP11-153D16
 Unpublished
 2 (bases 1 to 209462)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukigalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lenczky,J., Liu,C., Locke,K., MacDonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (18-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7321509.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1307
 Center clone name: 153_D_16
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 97% of reads
 Sequencing vector: Plasmid; n/a; %0-f% of reads
 3.29835082458771Chemistry: Dye-primer-amersham; 8% of reads
 Chemistry: Dye-terminator Big Dye; 92% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 172516 bases at least Q40
 Consensus quality: 188069 bases at least Q30
 Consensus quality: 195901 bases at least Q20
 Insert size: 175000; agarose-fp
 Insert size: 205562; sum-of-contigs
 Quality coverage

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 40 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1353: contig of 1353 bp in length
 * 1354 1453: gap of 100 bp
 * 1454 2778: contig of 1325 bp in length
 * 2779 2878: gap of 100 bp
 * 2879 4211: contig of 1333 bp in length
 * 4212 4311: gap of 100 bp
 * 4312 5622: contig of 1311 bp in length
 * 5623 5722: gap of 100 bp
 * 5723 7256: contig of 1534 bp in length
 * 7257 7357: gap of 100 bp
 * 7357 8828: contig of 1472 bp in length
 * 8829 10225: contig of 1297 bp in length
 * 10226 10325: gap of 100 bp
 * 10326 11376: contig of 1051 bp in length
 * 11377 11476: gap of 100 bp
 * 11477 12643: contig of 1167 bp in length

Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu

----- Project Information -----
 Center project name: M_BB0066C10

----- Summary Statistics -----

Sequencing vector: M13; 0%
 Chemistry: Dye-terminator; Plasmid, 100%
 Assembly program: Phrap; version 0.990319
 Consensus quality: 315826 bases at least Q40
 Consensus quality: 317232 bases at least Q30
 Consensus quality: 317973 bases at least Q20
 Insert size: 223000; agarose-fp
 Insert size: 322006; sum-of-contigs
 Quality coverage: 13.22 in Q20 bases; agarose-fp
 Quality coverage: 8.88 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1483: contig of 1483 bp in length
 * 1484 1583: gap of unknown length
 * 1584 3319: contig of 1736 bp in length
 * 3320 3419: gap of unknown length
 * 3420 4847: contig of 1428 bp in length
 * 4848 4947: gap of unknown length
 * 4948 8951: contig of 4004 bp in length
 * 8952 9051: gap of unknown length
 * 9052 13372: contig of 4321 bp in length
 * 13373 13472: gap of unknown length
 * 13473 28332: contig of 14860 bp in length
 * 28333 28432: gap of unknown length
 * 28433 64887: contig of 36455 bp in length
 * 64888 64987: gap of unknown length
 * 64988 158869: contig of 93882 bp in length
 * 158870 158969: gap of unknown length
 * 158970 321569: contig of 162600 bp in length
 * 321570 321669: gap of unknown length
 * 321670 322906: contig of 1237 bp in length.

FEATURES

Location/Qualifiers
 1. 322906
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clones="RP24-66C10"
 misc_feature
 1. 1483
 /note="assembly_name:Contig10"
 misc_feature
 1584. 3319
 /note="assembly_name:Contig12"
 misc_feature
 3420. 4847
 /note="assembly_name:Contig13"
 misc_feature
 4948. 8951
 /note="assembly_name:Contig15"
 misc_feature
 9052. 13372
 /note="assembly_name:Contig16"
 misc_feature
 13473. 28332
 /note="assembly_name:Contig17"
 misc_feature
 28433. 64887
 /note="assembly_name:Contig18"
 misc_feature
 64988. 158869
 /note="assembly_name:Contig19"
 misc_feature
 158970. 321569
 /note="assembly_name:Contig20"
 misc_feature
 321670. 322906
 /note="assembly_name:Contig7"

ORIGIN

Query Match 8.5%; Score 47.6; DB 2; Length 322906;
 Best Local Similarity 61.1%; Pred. No. 0.35;
 Matches 77; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 418 AAGAACAGGAGACAGAGACAAATTCAAATGCTCTTACACAGAGACTGATTATATAGATA 477
 Db 23449 AATAGCAAGAAGCTGGAAAGAACCCAGATGTCCTCAACAGAGAGATGGATACAGAAAT 23390
 QY 478 ATGGTACATTTCATTAACATGATATATTTACTTAATTAACAAACATACACACACACAACA 537
 Db 23389 ATGGTTCATTACACATGGAGAACTACTCAGCTATTAAACCAATGAATCTACAAAT 23330
 QY 538 CCTCGA 543
 Db 23329 CTTAGA 23324

RESULT 14

BX537253/c
 LOCUS BX537253
 DEFINITION Mouse DNA sequence from clone RP24-173B15 on chromosome 2, complete sequence.
 ACCSSION BX537253
 VERSION BX537253.6 GI:37518523
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 161021)
 AUTHORS Humphries, M.
 TITLE Direct Submission
 JOURNAL Submitted (02-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT On Oct 5, 2003 this sequence version replaced gi:32812728.
 Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> RP24-173B15 is from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

FEATURES

Location/Qualifiers
 1. 161021
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"

source

```

: /chromosome="2"
: /clone="RP24-173B15"
: /clone_lib="RPC1-24"
:
ORIGIN
:
Query Match      8.4%  Score 47.4  DB 10  Length 161021;
Best Local Similarity 54.2%  Pred. No. 0.43;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 356 ATGTATACATATTTAATACACAATGAGAACTCTTGGGGTGGTGGAGATGATAGGC 415
DB 69393 ATATATATATATATATACAGAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGA 69334
QY 416 TGAAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
DB 69333 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69274
QY 476 TAATGGTACATTCATATAAATGATATATTTACTTATTAATAACATACCAACACACA 532
DB 69273 TACTGGGAGAAAATGAAAACAAAACAAAACAAAACAAAACAAAACAAAACACA 69217

RESULT 15
AC133842
LOCUS      Rattus norvegicus clone CH230-185E3, *** SEQUENCING IN PROGRESS
DEFINITION      ***
ACCESSION      AC133842
VERSION        3  GI:30520933
KEYWORDS       HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE         Rattus norvegicus (Norway rat)
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 220394)
AUTHORS        Murny,D,Marie., Metker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
               Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
               Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
               Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
               Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
               Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
               Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
               Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
               Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
               Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
               Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
               Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duvai,B., Eaves,K.,
               Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
               Fraser,C.W., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Foster,P.,
               Georgiev,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
               Guaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
               Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
               Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
               Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
               Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
               Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
               Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
               Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
               Lorensheva,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,
               Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
               Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
               Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
               Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
               Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
               Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
               Nwackelmech,O., Okwunonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
               Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
               Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
               Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
               Reilly,B., Reilly,M., Reuter,Y., Reuter,M., Richards,S., Riggs,F.,
               Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFNZ
Center clone name: CH230-185E3
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 203828 bases at least Q40
Consensus quality: 208285 bases at least Q30
Consensus quality: 210314 bases at least Q20
Estimated insert size: 216328; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 220394: contig of 220394 bp in length.
* Location/Qualifiers
  1..220394
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-185E3"

```

```
misc_feature 1..3705
  /note="wgs_end_extension
  clone_end:T7"
  7885..8277
  /note="clone_boundary
  clone_end:T7
  site:EcoRI
  end_sequence:BH263108"
misc_feature
  complement(219195..219782)
  /note="clone_boundary
  clone_end:Sp6
  site:EcoRI
  end_sequence:BH263109"
ORIGIN
Query Match      8.3%; Score 47; DB 2; Length 220394;
Best Local Similarity 64.0%; Pred. No. 0.51;
Matches 71; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 418 AAGACAGGAGACAGACAAATTCAAATGCTTACACAGAGACTGATTATATAGATA 477
Db 45508 AATAGCCGAAATTTGTATACAAATTAGATGTTCTCAAGAGAGAGAAATAAATAAAT 45567
QY 478 ATGCTACATTTCATATAACATGATATATTTACTAATTAATAACATACCAACA 528
Db 45568 ATGCTACATTTCATATAACATGATATATTTACTAATTAATAACATACCAACA 528
```

Search completed: August 14, 2004, 19:01:42
Job time : 2448.69 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 1929.83 Seconds
(without alignments)
8711.839 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 ggracatcttggtgtggat.....gocgctagctcagtagctag 563

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_estl.*

10: gb_est2.*

11: gb_hc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.4	8.6	679	29	CE299099 tigr-gss-
C 2	48.2	8.6	535	10	BE635285 uv60a02.y
3	47.4	8.4	444	28	AZ003793 RPCI-23-3
4	46.6	8.3	459	28	AZ620342 LM0453u01

5	46	8.2	666	28	AQ933978
6	45.8	8.1	565	28	AQ933978
7	45.8	8.1	1201	9	AL553699
8	45.4	8.1	1101	29	CNS0106X
9	45	8.0	456	28	AQ556776
10	45	8.0	539	28	AZ101750
11	45	8.0	784	28	BZ160642
12	45	8.0	794	13	EX097137
13	45	8.0	897	28	BZ178903
14	45	8.0	3211	11	AK081359
15	44.4	7.9	345	28	BH069264
16	44.4	7.9	742	28	BZ092673
17	44.4	7.9	1201	9	AL548579
18	44.2	7.9	469	28	BH088169
19	44	7.8	365	28	AZ348426
20	43.8	7.8	1201	13	EX378989
21	43.4	7.7	184	28	AZ878343
22	43.4	7.7	370	9	AA966509
23	43.4	7.7	381	29	CE128378
24	43.4	7.7	388	28	AZ094991
25	43.4	7.7	464	28	AZ634360
26	43.4	7.7	492	28	AZ367086
27	43.4	7.7	499	9	AA926061
28	43.4	7.7	543	28	AZ873070
29	43.4	7.7	561	28	AZ278295
30	43.4	7.7	570	28	AZ284050
31	43.4	7.7	604	28	AQ923677
32	43.4	7.7	637	28	AQ923677
33	43.4	7.7	648	28	AZ376520
34	43.4	7.7	650	29	AG092150
35	43.4	7.7	796	28	AZ734214
36	43.4	7.7	897	28	BZ144279
37	43.2	7.7	514	28	AZ290946
38	43.2	7.7	535	28	AQ779549
39	43.2	7.7	574	28	BZ900086
40	43.2	7.7	600	28	AZ363255
41	43	7.6	782	28	BH042930
42	42.8	7.6	493	28	AZ431479
43	42.8	7.6	577	28	BH11528
44	42.8	7.6	617	28	AZ384488
45	42.8	7.6	629	28	AZ256692

ALIGNMENTS

RESULT 1
CE299099/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CE299099 679 bp DNA linear GSS 26-SEP-2003
tigr-gss-dog-17000359973606 Dog Library Canis familiaris genomic,
genomic survey sequence.

CE299099 GI:36086776

GSS.

Canis familiaris (dog)

Canis familiaris

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 679)

Kirkness,E.F., Baina,V., Halpern,A.L., Levy,S., Remington,K.,

Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and

Venter,J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

22875432

14512627

Contact: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirkness@tigr.org

and selected for ampicillin resistance."

```

Query Match      8.3%; Score 46.6; DB 28; Length 459;
Best Local Similarity 66.3%; Pred. No. 0.95;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY      418 AAGAACAGGAGACAGAGAACAATTCCAAATGTCTCTACACAGAGAAGACTGATTATAGATA 477
      |||
Ddb      300 AATAGCAGAGAGCTGAAACCAACCCAGATGTCTCTTACAGAGAAATTGATACAGAAATC 359
      |||
QY      478 ATGTACATTCATATAAAACATGATATATTACTATAATAAAA 518
      |||
Ddb      360 ATGTACATTTACACAATGAAGTACTATTTCAGCTATTAATA 400
      |||

RESULT 5
AQ333978
LOCUS      AQ333978
DEFINITION      RPCI-23-286P18.TV RPCI-23 Mus musculus genomic clone
ACCESSION      RPCI-23-286P18, genomic survey sequence.
VERSION      RPCI-23-286P18
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 666)
      Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S.,
      Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de
      Jong,P. and Fraser,C.M.
      Mouse BAC End Sequences from Library RPCI-23
      Unpublished (1999)
      Other GSSs: RPCI-23-286P18.TJ
      Contact: Shaying Zhao
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850, USA
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: szhao@tigr.org
      Clones are derived from the mouse BAC library RPCI-23. For BAC
      library availability, please contact Pieter de Jong
      (pieterdejong.med.buffalo.edu). Clones may be purchased from
      BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
      or from Resea cn Genetics (info@resgen.com). BAC end page:
      http://www.tigr.org/cdb/bac/ends/mouse/bac_end_intro.html
      Plate: 286 row: P column: 18
      Seq primer: T7
      Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..666
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-23-286P18"
                        /sex="Female"
                        /lab_host="DH10B"
                        /clone_lib="RPCI-23"
                        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                        brain genomic DNA was isolated and partially digested
                        with a combination of EcoRI and EcoRI Methylase. Size
                        selected DNA was cloned into the pBACe3.6 vector at the
                        EcoRI sites. The ligation products were transformed into
                        DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN
Query Match      8.2%; Score 46; DB 28; Length 666;
Best Local Similarity 60.3%; Pred. No. 1.3;
Matches 76; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```


BACN03K20 of DrosBAC library from *Drosophila melanogaster* (fruit fly); Genomic survey sequence.
AL098595
AL098595.1 GI:5610206
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Description of this BAC-end sequence was carried out as part of a collaboration with the European *Drosophila* Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This *Drosophila* melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosAcII.

FEATURES

source	location/Qualifiers
1..1101	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/clone="BACN03K20"
	/clone_lib="DrosBAC"
	/plasmid="pBelosAcII"
	/note="end : T7"

ORIGIN

```

Query Match      8.1%; Score 45.4; DB 29; Length 1101;
Best Local Similarity 19.1%; Pred.No.1.7;
Matches 99; Conservative 190; Mismatches 230; Indels 0; Gaps 0;

QY   26  TTGGACATCTTTTATTTCCTTTACCCTGTTCATCATCAAACTCTTAGAGTGTTACTTCTC 85
Db    TKKHCTCTCHKTITTKWTTTWKAKGTKTGTDDAKTKATMAIGTKTATRTWTAAKTA 637

QY   86  CAAGCGCGTTAACTTATTCAAGATTGACTTGAGTCTCCTATTTTCCTTTAAATTTACTAGA 145
Db    G::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: 637

QY   638 KRTADTKDTTAATAATAKATKKKARKDKDTAKDAKWMDKATKAKAKKAACAATAAAT 697
Db    A::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   146 AGTGAGCTCCAAAGAACTACAGAAAATAGAGGAAGTCTCCAATTGAGCCATGAATGTG 205
Db    A::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   698 ATGKDARAKAKAKATKKAKARAMDTATWTATADAADKGAKDKAKAKAKADARDKR 757
Db    A::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   206 AGCACCTGGCATTTAAGCATGAAGATGAGGCTTCTATGTTAGGGACTGGAGTAGCAGC 265
Db    RWDAKDKRKRKAACAADDADADAKADAKADADAADDDGRGDGKKRAKDRK 817

QY   266 ATTCCAGGAAGATCTCAGAGTCGAAACAATAGATTATCAGTTTAATCTCTTGGAC 325
Db    :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   818 KKXKKDKAWGDKKKKAKDKAAAKAKADADGAKAKERRAGDKDKAKADAKAAKAKAKDD 877
Db    :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   326 CAAAGAGAGCCTTGAATTCCTGGCTCGGTGATGTATACATATTTAATACCAATGAGAG 385
Db    DDAAKATKAKKATKAKKDKAKKKKKDKDKAKAKAKAKADAKDDKDDKDDKAKA 937

QY   386 CTCCTTGGGGTGAAGTGAAGTGAAGGCTGAAGAACAGGAGACAGAGAACAATTCAAA 445
Db    :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   938 DKKKWKDKRAKDKKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 997
Db    :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY   446 TGTCCTTTACACAGAACACTGATTATATAGATATGTTACATTCATTAACATGATATAT 505
Db    DDAADAKAKADAKAKAKADDAKAKAKADAKADAKADAKADAKAKAKAKAKADA 1057

QY   506 TTACTAATTAACATACCAACACACACACACCTCGAG 544
Db    :| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

```

RESULT 10
AZ101750
LOCUS
DEFINITION
RPCI-23-452H21.TV RPCI-23 Mus musculus genomic clone
RPCI-23-452H21, genomic survey sequence.
ACCESSION
AZ101750
VERSION
AZ101750.1 GI:7754806
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 539)
Zhao S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S.,
Akinret, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de
Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library RPCI-23
Mouse BAC End Sequences from Library RPCI-23
Other GSSs: RPCI-23-452H21.TJ
Unpublished (1999)
Contact: Shaving Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 452 row: H column: 21
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..539
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-452H21"
/sex="Female"
/lab_host="PH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN
Query Match 8.0%; Score 45; DB 28; Length 539;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 75; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 418 AAGAACAGGACAGACAGCAATTCCTTACACAGACGACTGATTATATAGATA 477
DB 308 ATATGCCAGAGCTGGAAAGACCAGGATGTCCTTCAACAGGAGGATGGATCAAAAAT 367
QY 478 ATGGTACATTCATTAACATGATATATTTACTAATTAACATCAACACACACACA 537
DB 368 GTGGTACATTTACAAATGGAGTACTATTCAGCTATTAAAGAAAGTGAATACACAAAAT 427
QY 538 CCTCG 542
DB 428 CCTAG 432

RESULT 11
BZ160642

LOCUS
DEFINITION
CH230-327L4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-327L4, genomic survey sequence.
ACCESSION
BZ160642
VERSION
BZ160642.1 GI:23801598
KEYWORDS
GSS.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
REFERENCE
1 (bases 1 to 784)
Zhao S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, P., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-327L4.TJ
Contact: Shaving Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 327 row: L column: 4
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..784
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-327L4"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pIRBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
Pieter de Jong"
ORIGIN
Query Match 8.0%; Score 45; DB 28; Length 784;
Best Local Similarity 65.3%; Pred. No. 2.2;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 418 AAGAACAGGACAGACAGCAATTCCTTACACAGACGACTGATTATATAGATA 477
DB 499 AATACCCAGAAATGGAAACAATCGGATGTCCTTACACAGACGACTGGATACATAAAT 558
QY 478 ATGGTACATTCATTAACATGATATATTTACTAATTTAAAA 518
DB 559 GTGGTACATTTACCAATGGAGTACTACTCAACTATTAAAA 599

RESULT 12
BX097137/c
LOCUS
DEFINITION
BX097137 Soares multiple sclerosis 2NBHMS Homo sapiens cDNA clone
IMAGD98804619, IMAGE:280131, mRNA sequence.
ACCESSION
BX097137
VERSION
BX097137.1 GI:27843097
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
	Genome Res. 10 (10), 1617-1630 (2000)	
	20499374	
JOURNAL MEDLINE PUBLISHED	11042159	
	3	
	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, F., Nakamura, S., Hazama, M., Nashine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL MEDLINE PUBLISHED	Genome Res. 10 (11), 1757-1771 (2000)	
	20530913	
	11076861	
	4	
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
	Functional annotation of a full-length mouse cDNA collection	
	Nature 409, 685-690 (2001)	
	5	
REFERENCE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
	Nature 420, 563-573 (2002)	
	6 (bases 1 to 3211)	
JOURNAL MEDLINE PUBLISHED	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
	Direct Submission	
	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp].	
	URL: http://genome.gsc.riken.go.jp/ , Tel: 81-45-503-9222, Fax: 81-45-503-9216	
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.	
	Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
	Please visit our web site for further details.	
	URL: http://genome.gsc.riken.go.jp/	
FEATURES	URL: http://fantom.gsc.riken.go.jp/	
	Location/Qualifiers	
	1. .3211	
	/organism="Mus musculus"	
source	/mol_type="mRNA"	
	/strain="C57BL/6J"	
	/db_xref="FANTOM DB:C130010D01"	
	/db_xref="MGI:2413724"	
misc_feature	/db_xref="taxon:10090"	
	/clone="C130010D01"	
	/tissue type="head"	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
misc_feature	/dev_stage="16 days embryo"	
	1. .3211	
	misc_feature	
	misc_feature	
ORIGIN	Query Match	
	Best Local Similarity 8.0%; Score 45; DB 11; Length 3211;	
	Matches 75; Conservative 0; Mismatches 50; Indels 0; Gaps 0;	
	QY 418 AAGAACAGGAGACAGACAGCAATTCAAATGTCCTTACAGACAGACATGATTATACATA 477	
ORIGIN	Db 2309 AATAGCCAGAAAGCTGGAAAGAACCCAGATGTCCTCAACAGAAATGAATATAAAAAAT 2368	
	QY 478 ATGGTACATTCATATAACATGATATATTACTAATTAATAAATACACACACACACAACA 537	
	Db 2369 ATGGTACATTCATATAACATGATATATTACTAATTAATAAATACACACATGAATTCATGAATT 2428	
	QY 538 CCTCG 542	
ORIGIN	Db 2429 CTTAG 2433	
	RESULT 15	
	BH069264	
LOCUS	PCPI-24-257C9.TJ PCPI-24 Mus musculus genomic clone PCPI-24-257C9,	
	DEFINITION	
	Genomic survey sequence.	
	ACCESSION	
VERSION	BH069264.1 GI:14888861	
	KEYWORDS	
	GSS.	
	SOURCE	
ORGANISM	Mus musculus (house mouse)	
	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 345)	
	Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, W., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.	
	Mouse BAC End Sequences from Library PCPI-24	
	Unpublished (1999)	
JOURNAL	Contact: Shaving Zhao	
	Department of Eukaryotic Genomics	
	The Institute for Genomic Research	
	9712 Medical Center Dr., Rockville, MD 20850, USA	
COMMENT	Tel: 301 838 0200	
	Fax: 301 838 0208	
	Email: szhao@tigr.org	
	Clones are derived from the mouse BAC library PCPI-24. For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html	
FEATURES	Plate: 257 row: C column: 9	
	Seq primer: SP6	
	Class: BAC ends.	
	Location/Qualifiers	
source	1. .345	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/strain="C57BL/6J"	
misc_feature	/db_xref="taxon:10090"	
	/clone="PCPI-24-257C9"	
	/sex="Male"	
	/cell_type="Spleen/Brain"	
ORIGIN	/clone_lib="PCPI-24"	
	/notes="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; PCPI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."	
	Query Match	
	7.9%; Score 44.4; DB 28; Length 345;	

Best Local Similarity 59.5%; Pred. NO. 3.3;
Matches 75; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY	418	AAGAAACAGGACAGACAGACAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATA	477
Db	16	AATAGCCAGAGCTGGAAACAAACCCCAATGTCCCTCACACAGAGAAATCAATACAGAAAT	75
QY	478	ATGGTACATTCATATAAACATGATATATTTACTAATTHAAACATACCAACACACACAACA	537
Db	76	ATGGTACATTTACACAATGGAGTACTACTCAGCAATTAACAAACAAATGAATTCATGAAAT	135
QY	538	CCTCGA	543
Db	136	CTTAGA	141

Search completed: August 14, 2004, 20:35:28
Job time : 1935.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 283.474 Seconds
8437.229 Million cell updates/sec

Title: US-10-082-830-99
Perfect score: 563
Sequence: 1 ggtacattctggctgtggat.....gccgctagctcagtagctag 563

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseq1380s.*
2: Geneseq1930s.*
3: Geneseq2000s.*
4: Geneseq2001bs.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	563	100.0	563	6	ABK93493 Human bre
2	43.4	7.7	75798	7	ABX77212 Mouse uri
3	42.4	7.5	445	4	AAI87875 Human pol
4	42.4	7.5	7374	4	AAI07519 Human rep
5	42.4	7.5	7374	4	ABA08211 Human ova
6	42	7.5	16766	6	ABL34157 Human imm
7	42	7.5	144792	9	ADC87620 Human GPC
8	42	7.5	349981	9	ADC87619 Human GPC
9	41.2	7.3	167343	6	ABL64403 Stomach c
10	41.2	7.3	167343	6	ABL67239 Thyroid c
11	40.8	7.2	77425	6	ABK83502 Human cDN
12	40.2	7.1	17032	6	AAK84885 Human imm
13	40.2	7.1	173810	6	ABN85752 Mouse chr
14	40	7.1	11027	4	AAI37290 Human mus
15	40	7.1	11027	4	AAI37289 Human mus
16	40	7.1	11027	7	ABX60277 cDNA enco
17	40	7.1	11027	7	ABX60278 cDNA enco
18	40	7.1	104000	8	AAI57669 Human pho
19	39.6	7.0	21837	4	AAK85946 Human imm
20	39.4	7.0	323	4	AAI36295 Human mus
21	39.4	7.0	323	4	AAI36293 Human mus
22	39.4	7.0	323	7	ABX59283 cDNA enco
23	39.4	7.0	323	7	ABX59281 cDNA enco

C	24	39.4	7.0	1771	5	AAH22966 Human pho
	25	39.4	7.0	2890	4	AAH77031 Human NF-
	26	39.4	7.0	110000	5	AAF84800 Nucleotid
C	27	39.2	7.0	9725	6	ABL33292 Human imm
C	28	39.2	7.0	9725	6	ABN80180 Human che
	29	39.2	7.0	21537	6	ABL33999 Human imm
C	30	39	6.9	606	4	AAK63763 Human imm
C	31	39	6.9	4527	4	AAK85875 Human imm
C	32	39	6.9	81001	4	AAF30035 Human apo
C	33	38.8	6.9	179	2	AAI19869 Human gen
C	34	38.8	6.9	512	4	AAI186036 Human pol
C	35	38.8	6.9	9507	7	AAI07097 Human rep
C	36	38.8	6.9	9507	7	ABZ67483 Human sec
C	37	38.8	6.9	9507	7	ABZ73900 Secrete
	38	38.6	6.9	768	4	AAK86425 Human imm
	39	38.6	6.9	768	4	AAK86423 Human imm
	40	38.6	6.9	768	4	AAK86426 Human imm
	41	38.6	6.9	16766	6	ABL34156 Human imm
C	42	38.6	6.9	173810	6	ABN85752 Mouse chr
C	43	38.4	6.8	14919	4	AAI46506 Tumour su
	44	38.4	6.8	33923	4	AAK67071 Human imm
	45	38.4	6.8	149671	6	ABK84797 Human cDN

ALIGNMENTS

RESULT 1
ABK93493
ID ABK93493 standard; cDNA; 563 BP.
XX
AC ABK93493;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human breast specific nucleic acid, BSNA #99.
XX
KW Human; ss; breast specific nucleic acid; BSNA; breast cancer;
KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
XX
OS Homo sapiens.
XX
PN WO200236807-A2.
XX
PD 10-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-US046888.
XX
PR 27-OCT-2000; 2000US-0243802P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
XX
DR WPI; 2002-463415/49.
XX
PT New breast-specific nucleic acids and polypeptides, useful for
PT identifying, diagnosing, monitoring, staging, imaging, and treating
PT breast cancer and non-cancerous disease states in breast tissues.
XX
PS Claim 1; Page 191-192; 281pp; English.
XX
CC The invention relates to breast specific nucleic acids (BSNA) and breast-
CC specific polypeptides (BSP). Also included are a method for determining
CC the BSNA in a sample, a vector comprising a BSNA, a host cell comprising
CC the vector, a method for producing a polypeptide encoded by a BSNA, an
CC anti-BSP antibody and a method for determining the presence of a BSP in a
CC sample. The breast-specific nucleic acids, polypeptides and compositions
CC comprising them are useful for identifying, diagnosing, monitoring,
CC staging, imaging, and treating breast cancer, mammary tumour and non-
CC cancerous disease states in breast tissue; for identifying breast tissue;
CC for monitoring, identifying and/or designing agonists and antagonists of
CC the polypeptides; in gene therapy; in producing transgenic animals and

CC	cells; for producing engineered breast tissue for treatment and research;	PN	WO200283897-A1.
CC	and as elements in an array or computer program for pattern recognition	XX	
CC	of breast disorders. The nucleic acids may be used as hybridisation	PD	24-OCT-2002.
CC	probes to detect, characterise and quantify hybridising nucleic acids in,	XX	
CC	and isolate hybridising nucleic acids from, both genomic and transcript-	PF	18-APR-2002; 2002WO-AU000485.
CC	derived nucleic acid samples. The BSP protein may be used in a vaccine	XX	
CC	composition for raising an immune response against breast cancer. The	PR	18-APR-2001; 2001AU-00004467.
CC	present sequence is BSN cDNA of the invention	XX	(GENE-) GENE STREAM PTY LTD.
XX		PA	
XX		XX	Daly JM;
SQ	Sequence 563 BP; 191 A; 99 C; 118 G; 155 T; 0 U; 0 Other;	PI	
	Query Match 100.0%; Score 563; DB 6; Length 563;	XX	
	Best Local Similarity 100.0%; Pred. No. 6.1e-141;	DR	WPI; 2003-093021/08.
	Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	
QY	1 GGTACATCTGGCTGGTGGATGGAAATTTGACATACCTTTTATTTCTTACCTGTTACAT 60	PT	New transgenic non-human animal expressing a foreign polypeptide
Db	1 GGTACATCTGGCTGGTGGATGGAAATTTGACATACCTTTTATTTCTTACCTGTTACAT 60	PT	associated with drug behavior and/or metabolism, useful for studying the
		PT	behavior and/or metabolism of a drug in other animals.
QY	61 ATCAAACTTAGGATGATTACTTCCAAAGCGGTAACTTATTCAGATTGACTGATC 120	PS	Disclosure; Page 354-394; 408pp; English.
Db	61 ATCAAACTTAGGATGATTACTTCCAAAGCGGTAACTTATTCAGATTGACTGATC 120	XX	
QY	121 TCCTATTTTCTTAAATTTACTAGAAGTGAGGCTCCAAAGAACTACAGAAATAGAGGAA 180	CC	This invention relates to a transgenic non-human animal which may be used
Db	121 TCCTATTTTCTTAAATTTACTAGAAGTGAGGCTCCAAAGAACTACAGAAATAGAGGAA 180	CC	for assessing the behaviour and/or metabolism of a drug in another animal
QY	181 AGTCTCCATTGAGCCATGAACTGTGACACCTGGCATTTAAGCATGAAGAGTAGGGCTTC 240	CC	and which expresses a foreign polypeptide associated with drug behaviour
Db	181 AGTCTCCATTGAGCCATGAACTGTGACACCTGGCATTTAAGCATGAAGAGTAGGGCTTC 240	CC	and/or metabolism. The invention also comprises a nucleic acid construct
QY	241 TATGTTAGGAGTGGAGTAGGAGCATTCCAGGAAAGGATCTCAGAGGTCAGAAACAATA 300	CC	for use in producing the above transgenic non-human animal and a method
Db	241 TATGTTAGGAGTGGAGTAGGAGCATTCCAGGAAAGGATCTCAGAGGTCAGAAACAATA 300	CC	of assessing the metabolism and/or behavior of a drug in an animal of
QY	301 GATTATCAGTTAAATACCTTCTGGACCAAGAGACCTTGAATCCCTGGCTCGGTGATGTA 360	CC	interest, comprising administering a test agent to the transgenic animal
Db	301 GATTATCAGTTAAATACCTTCTGGACCAAGAGACCTTGAATCCCTGGCTCGGTGATGTA 360	CC	and conducting analytical tests to determine drug metabolism and/or
QY	361 TACATATTTAATACAAATGAGAGCTCCCTTGAGGGTGAGTGAGAGTGATAGGCTGAAG 420	CC	behaviour. The transgenic animal is useful in studying drug metabolism
Db	361 TACATATTTAATACAAATGAGAGCTCCCTTGAGGGTGAGTGAGAGTGATAGGCTGAAG 420	CC	and/or behaviour in other animals. The nucleic acid construct is useful
QY	421 AACAGGAGACAGAGACAATTCAAATGCTCTTACAGAGAGACCTGATTATATAGATAATG 480	CC	in producing the above transgenic animal and the methods are used for
Db	421 AACAGGAGACAGAGACAATTCAAATGCTCTTACAGAGAGACCTGATTATATAGATAATG 480	CC	producing, breeding and using transgenic animals for pharmacological
QY	481 GTACATTCATATAACATGATATATTTACTAATTTAAACATACCACACACACACACCT 540	CC	(e.g. pharmacokinetic or pharmacodynamic assays) and/or toxicological
Db	481 GTACATTCATATAACATGATATATTTACTAATTTAAACATACCACACACACACACCT 540	CC	studies. Nucleic acid sequences used within the invention are serum
QY	541 CGAGCCGCTAGTCTCGAGTCTAG 563	CC	albumin; alpha-acidic glycoprotein; cytochrome P450 (CYP); uridine
Db	541 CGAGCCGCTAGTCTCGAGTCTAG 563	CC	diphosphoglucuronosyl transferase (UGT); multidrug resistance proteins
		CC	and (MAP's). The present sequence represents a PCR primer used to create
		CC	a transgenic animal within the scope of the invention
		XX	
		SQ	Sequence 75798 BP; 21534 A; 17028 C; 16619 G; 20617 T; 0 U; 0 Other;
			Query Match 7.7%; Score 43.4; DB 7; Length 75798;
			Best Local Similarity 59.2%; Pred. No. 0.6;
			Matches 74; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY	418 AAGAACAGAGACAGACAACTTCAATGTCCTTACACAGAGACTGATTATAGATA 477		
Db	35651 AATAGCTAGAAGCTGGAAGAACCCAGATGTCCTTCAACAGAGGAATGGATACAAAATT 35710		
QY	478 ATGGTACATTCATATAAACAATGATATATTTACTAATTTAAACATACCACACACACA 537		
Db	35711 GTGGTACATTTACACAATGGAGTACTCTCAGCTATTATAAAGAATGAATTCACAAAATT 35770		
QY	538 CCTCG 542		
Db	35771 CCTAG 35775		
			RESULT 3
			AA187875
			ID AA187875 standard; cDNA; 445 BP.
			XX AA187875;
			AC AA187875;
			XX AA187875;
			DT 06-NOV-2001 (first entry)
			XX Human polynucleotide SEQ ID NO 7935.
			DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
			XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
			KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
			KW tissue growth factor; immunomodulatory; cancer; leukaemia;
			XX uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.
			OS Homo sapiens.
			XX

```
OS Homo sapiens.
XX WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX PR 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
XX PA (HYSE-) HYSEQ INC.
XX FI Tang YT, Liu C, Dmanac RT;
XX DR WPI; 2001-514838/56.
XX DR P-PSDB; AAO07944.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX PS Claim 1; SEQ ID NO 7935; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 445 BP; 158 A; 75 C; 109 G; 103 T; 0 U; 0 Other;

Query Match 7.5%; Score 42.4; DB 4; Length 445;
Best Local Similarity 62.0%; Pred. No. 0.22;
Matches 67; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 430 CAGAGACAAATTCAAATGTCCTTACACAGAGACTGATTATATAGATAATGGTACATTCA 489
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 CTGGAACAACTCAATGTTCTTAAACAGGAGATGATGAAGTTTATGGTATTCA 207

QY 490 TATAACATGATATATTACTAATTAACATACCAACACACACACACA 537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 CATATAGAACTACTCTGACATAAAGAGATAAAGTCTCTGGAACA 255

RESULT 4
AAL07519/c
ID AAL07519 standard; DNA; 7374 BP.
XX AC AAL07519;
XX XX
XX DT 21-NOV-2001 (first entry)
XX XX
XX DE Human reproductive system related antigen DNA SEQ ID NO: 10207.
XX XX
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX OS Homo sapiens.
XX XX
XX PN WO200155320-A2.
XX XX
XX PD 02-AUG-2001.
XX XX
XX PF 17-JAN-2001; 2001WO-US0011339.
XX XX
```


Query Match 7.5%; Score 42.4; DB 4; Length 7374;
 Best Local Similarity 57.6%; Pred. No. 0.53;
 Matches 76; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 387 TCCTTGAGGTGAGTGTAGGCTGGAAGACAGAGACAGACAAATCAAT 446
 Db 1091 TACATGAATGTTGTTGAAGCATTATACATAATGACCAAAAAATGAAATTAATTCAAAT 1032

QY 447 GTCCTTACAGAGAGCTGATTATATAGATAAATGGTACATTCATATAAACAATGATATATT 506
 Db 1031 GTCCTTACAGAGAGCTGATTATATAGATAAATGGTACATTCATATAAACAATGATATATT 506

QY 507 TACTAATTAATAA 518
 Db 971 CAGCAATAAAAA 960

RESULT 6
 ABL34157/c
 ID ABL34157 standard; DNA; 16766 BP.
 AC ABL34157;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 2130.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; anidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2002-130909/17.
 DR
 PT Nucleic acid' comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 2130; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 16766 BP; 4604 A; 267 C; 3632 G; 8263 T; 0 U; 0 Other;

Query Match 7.5%; Score 42; DB 6; Length 16766;
 Best Local Similarity 60.5%; Pred. No. 0.88;

Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 422 ACAGGACAGACAGACAAATTCATTCCTTACACAGAACTGATATATAGATATGG 481
 Db 503 AAAAAAATAAATAATTCATTCCTTACACAGAACTGATATATAGATATGG 444

QY 482 TACATTCATTAACATGATATATTTACTAATTAATAACATACCAACACACACAA 535
 Db 443 TATATCTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 390

RESULT 7
 ADC87620/c
 ID ADC87620 standard; DNA; 144792 BP.
 AC ADC87620;
 XX
 DT 01-JAN-2004 (first entry)
 DE Human GPCR related polynucleotide SEQ ID NO:2073.
 XX
 XX ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1270724-A2.
 XX
 PD 02-JAN-2003.
 XX
 PF 18-JUN-2002; 2002EP-00013517.
 XX
 PR 18-JUN-2001; 2001JP-00246789.
 XX
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX
 PI Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX
 XX WPI; 2003-315783/31.
 DR
 PT New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX
 PS Disclosure; SEQ ID NO 2073; 28pp; English.
 XX
 CC The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC87618-ADC87623 represent polynucleotide sequences
 CC related to the invention.
 XX
 SQ Sequence 144792 BP; 39827 A; 32142 C; 33413 G; 39310 T; 0 U; 100 Other;

Query Match 7.5%; Score 42; DB 9; Length 144792;
 Best Local Similarity 66.7%; Pred. No. 1.7;
 Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 426 GACACAGACAGACAAATTCATTCCTTACACAGAACTGATATATAGATATGGTACA 485
 Db 9727 GAAAGAGAGAGCAATTCATTCCTTACACAGAACTGATATATAGATATGGTACA 9668

QY 486 TTCATATAAACAATGATATATTTACTAATTA 515
 Db 9667 TGCATCTATTCATGAGATATTTAGATTTA 9638

RESULT 8
 ADC87619/c

ID ADC87619 standard; DNA; 349981 BP.
 AC ADC87619;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 DE Human GPCR related polynucleotide SEQ ID NO:2072.
 XX
 KW ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 KW gene therapy.
 OS Homo sapiens.
 XX
 PN BP1270724-A2.
 XX
 PD 02-JAN-2003.
 XX
 PF 18-JUN-2002; 2002EP-00013517.
 XX
 PR 18-JUN-2001; 2001JP-00246789.
 XX
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX
 PI Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX
 DR WPI; 2003-315783/31.
 XX
 PT New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX
 PS Disclosure; SEQ ID NO 2072; 28pp; English.
 XX
 CC The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC87618-ADC87623 represent polynucleotide sequences
 CC related to the invention.
 XX
 SQ Sequence 349981 BP; 100900 A; 78560 C; 77396 G; 92823 T; 0 U; 302 Other;
 Query Match 7.5%; Score 42; DB 9; Length 349981;
 Best Local Similarity 66.7%; Pred. No. 2,3;
 Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 QY 426 GAGACAGAGACAATCTCAATCTCTTACACAGAGACTGATTATATAGATAATGGTACA 485
 DB 309728 GAAAGAGGAGCAATCTGATCTCGTCAATAGGGGACTGTTGGATAAATAATGCTGCT 309669
 QY 486 TTCATATAACATGATATATTACTAATTA 515
 DB 309668 TGCATCTATTCTAGATATTTTAGATTTA 309639
 RESULT 9
 ABL64403
 ID ABL64403 standard; DNA; 167343 BP.
 XX
 AC ABL64403;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Stomach cancer related gene sequence SEQ ID NO:2740.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX

OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 PR 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233113P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 22-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237284P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 01-NOV-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Hortigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-189264/24.
 XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 2740; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in CC expression is indicative of anti-neoplastic activity. (I) has cytostatic CC activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is

CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1 and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 0 U; 13 Other;
 Query Match 7.3%; Score 41.2; DB 6; Length 167343;
 Best Local Similarity 67.4%; Pred. No. 3;
 Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
 QY 409 ATAGGCGCTGAAGACAGGAGACAGACAAATTCAAATGCTCTACACAGAGACTGATT 468
 Db 159637 ATTGTTCAAAATACAAAGATTGGAAGAGGCAATATCTTGAGTAGAGACTGATG 159696
 QY 469 ATATAGATTAATGGTACATTCATATAA 494
 Db 159697 AAATACATTGTGCTACATACATACAA 159722
 RESULT 10
 ABL67239
 ID ABL67239 standard; DNA; 167343 BP.
 XX
 AC ABL67239;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Thyroid cancer related gene sequence SEQ ID NO:5576.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200194629-A2.
 XX
 XX 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 PR 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0231313P.
 PR 18-SEP-2000; 2000US-0231313P.
 PR 20-SEP-2000; 2000US-0233611P.
 PR 20-SEP-2000; 2000US-0234003P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236851P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237435P.
 PR 03-OCT-2000; 2000US-0237588P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI; 2002-188264/24.
 XX
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 PS Claim 1; SEQ ID NO 5576; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 0 U; 13 Other;
 Query Match 7.3%; Score 41.2; DB 6; Length 167343;
 Best Local Similarity 67.4%; Pred. No. 3;
 Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
 QY 409 ATAGGCGCTGAAGACAGGAGACAGACAAATTCAAATGCTCTACACAGAGACTGATT 468
 Db 159637 ATTGTTCAAAATACAAAGATTGGAAGAGGCAATATCTTGAGTAGAGACTGATG 159696
 QY 469 ATATAGATTAATGGTACATTCATATAA 494
 Db 159697 AAATACATTGTGCTACATACATACAA 159722
 RESULT 11
 ABL67239/c
 ID ABL67239 standard; CDNA; 77425 BP.
 XX
 AC ABL67239;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #73.

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US030821.
XX
XX 03-OCT-2000; 2000US-0237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
XX genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.
XX
XX Claim 1; SEQ ID NO 73; 11app; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing the
XX expression level to an expression level in an unactivated GC, where
XX differential expression of Gs is indicative of GCA. Also included are
XX modulating (M2) GA by contacting GC with an agent that alters the
XX expression of at least one gene in Gs; (2) screening (M3) for an agent
XX capable of modulating GCA or an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease using the gene expression
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease, by detecting the level of
XX expression of the gene is indicative of inflammation; (4) treating
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease, by contacting a tissue having inflammation with an
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful
XX for screening an agent capable of modulating GCA preferably in an
XX inflammation in a tissue; M4 is useful for detecting an inflammation
XX (especially chronic) in a tissue, an allergic response in a subject,
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, adult
XX respiratory distress syndrome, inflammatory bowel disease, Crohn's
XX disease, ulcerative colitis, periodontal disease; also bacterial
XX infection, viral infection, parasitic infection, protozoal infection,
XX fungal infection and M5 is useful for treating one of the above
XX conditions. The present sequence represents a gene differentially
XX expressed in granulocytes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 77425 BP; 23611 A; 13793 C; 14665 G; 25356 T; 0 U; 0 Other;
Query Match 7.2%; Score 40.8; DB 6; Length 77425;
Best Local Similarity 59.5%; Pred. No. 3;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 417 GAAGAAACAGGAGACAGAGCAATTCCTTACACAGAGACTGATTATATAGAT 476
Db 39110 GCATAGCCTCAAACTAGAAACAATCCAAATGTCCTAAGACAGAGCTAACAAATAAT 39051
QY 477 AATGGTACATTCATATAAATCATATATTTACTTAATTAACATACCAACACACA 532
Db 39050 GGTGGTATATTCATATATATACCTTATATCTATGGAATATAACATGAGAATGCATA 38995
RESULT 12
AAK84885
ID AAK84885 standard; DNA; 17032 BP.
XX
XX AAK84885;
AC
XX
XX 07-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39697.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
XX 17-MAR-2000; 2000US-0190076P.
PR
XX 18-APR-2000; 2000US-0198123P.
PR
XX 19-MAY-2000; 2000US-0205515P.
PR
XX 07-JUN-2000; 2000US-0209467P.
PR
XX 28-JUN-2000; 2000US-0214886P.
PR
XX 30-JUN-2000; 2000US-0215135P.
PR
XX 07-JUL-2000; 2000US-0216647P.
PR
XX 07-JUL-2000; 2000US-0216880P.
PR
XX 11-JUL-2000; 2000US-0217487P.
PR
XX 14-JUL-2000; 2000US-0217496P.
PR
XX 26-JUL-2000; 2000US-0218290P.
PR
XX 26-JUL-2000; 2000US-0220963P.
PR
XX 14-AUG-2000; 2000US-0224518P.
PR
XX 14-AUG-2000; 2000US-0224519P.
PR
XX 14-AUG-2000; 2000US-0225213P.
PR
XX 14-AUG-2000; 2000US-0225214P.
PR
XX 14-AUG-2000; 2000US-0225266P.
PR
XX 14-AUG-2000; 2000US-0225267P.
PR
XX 14-AUG-2000; 2000US-0225268P.
PR
XX 14-AUG-2000; 2000US-0225270P.
PR
XX 14-AUG-2000; 2000US-0225447P.
PR
XX 14-AUG-2000; 2000US-0225575P.
PR
XX 14-AUG-2000; 2000US-0225758P.
PR
XX 18-AUG-2000; 2000US-0226279P.
PR
XX 22-AUG-2000; 2000US-0226681P.
PR
XX 22-AUG-2000; 2000US-0226868P.
PR
XX 23-AUG-2000; 2000US-0227182P.
PR
XX 30-AUG-2000; 2000US-0227009P.
PR
XX 01-SEP-2000; 2000US-0228924P.
PR
XX 01-SEP-2000; 2000US-0229287P.
PR
XX 01-SEP-2000; 2000US-0229343P.
PR
XX 01-SEP-2000; 2000US-0229344P.
PR
XX 05-SEP-2000; 2000US-0229345P.
PR
XX 05-SEP-2000; 2000US-0229509P.
PR
XX 05-SEP-2000; 2000US-0229513P.
PR
XX 06-SEP-2000; 2000US-0230437P.

DE Mouse chromosome 11 BAC clone RP23-270L8 SEQ ID NO 10.
XX Mouse; Can 1; antifertility; gynaecological; infertility;
KW premature ovarian failure; menopause; Sertoli Cell only syndrome;
KW BAC clone RP23-270L8; GenBank reference AC083815; chromosome 11; ds.
XX
OS Mus musculus.
XX
PN US2002119929-A1.
XX
PD 29-AUG-2002.
XX
PF 02-NOV-2001; 2001US-00003806.
XX
PR 03-NOV-2000; 2000US-0245872P.
XX
PA (BISH/) BISHOP C E.
PA (AGOU/) AGOUNIK A I.
PA (ZHUQ/) ZHU Q.
XX
PI Bishop CE, Agounik AI, Zhu Q;
XX
DR WPI; 2002-618953/66.
XX
PT A nucleic acid molecule (I) encoding a Can 1 polypeptide used in treating
PT infertility.
XX
PS Disclosure; Page; 45pp; English.
XX
CC The invention relates to a nucleic acid molecule (I) encoding a Can 1
CC polypeptide. The Can 1 nucleic acid molecule is used to diagnose or treat
CC infertility or premature ovarian failure or Sertoli Cell only syndrome in
CC a mammal. The present sequence is that of a mouse Can 1 encoding
CC chromosome 11 BAC clone RP23-270L8 of the invention. Note: The present
CC sequence is not given in the printed specification but was obtained
CC through the GenBank reference AC083815
XX
SQ Sequence 173810 BP; 50474 A; 33519 C; 33994 G; 55823 T; 0 U; 0 Other;

Query Match 7.1%; Score 40.2; DB 6; Length 173810;
Best Local Similarity 57.6%; Pred. No. 5.6;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGACAGACAAATCAATGTCCTTACACAGAGACTGATTATATAGATA 477
Db 42876 AATAGCCAGACCTGGAAACACCAGATGTCCTTTAAACAGGATAGATACAGAAAT 42935

QY 478 ATGTCATTCATATAAACAATGATATATTTACTAATTAACACATACCACACACACAACA 537
Db 42936 ATGTCATTCATATAAACAATGATATATTTACTAATTAACACATACCACACACACAACA 537

QY 538 CCTCG 542
Db 42996 CTTAG 43000

RESULT 14
AAL37290
ID AAL37290 standard; DNA; 11027 BP.
XX
AC AAL37290;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3655.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX

OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001338.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189974P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.

```

PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal
PT cancers and also for testing and detection e.g. diagnosis.
XX Example 2; SEQ ID NO 3655; 781bp + Sequence Listing; English.
XX The invention relates to novel genes (AAL37289-AAL37666) and proteins
CC (ABR03087-ABR04109) associated with the musculoskeletal system useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 11027 BP; 2851 A; 2179 C; 2539 G; 3458 T; 0 U; 0 Other;
Query Match 7.1%; Score 40; DB 4; Length 11027;
Best Local Similarity 55.9%; Pred. No. 2.7;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 417 GAAGAACAGGAGACAGACAAATTCATTCCTTACACAGACAGCTGATTATATAGAT 476
DB 5908 GAGGAACAAAAGCTTAAAGACATCAAAATGCTCTTTAATAGAGACTGTCTTAATCAAT 5967
QY 477 AATGTCATATTCATATAACATGATATATTACTAATTAATAACATACCAACACACACAAC 536
DB 5968 TAGGCATATTCATTCAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 6027
QY 537 ACCTCGAGCGCTAGT 552
DB 6028 TGAGGAGGCTTTCAGT 6043
RESULT 15
AAL37289
ID AAL37289 standard; DNA; 11027 BP.
XX AAL37289;
XX 08-JAN-2002 (first entry)
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3654.
DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
XX Homo sapiens.
XX WO200155367-A1.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001338.
```

XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225470P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226661P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227709P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-023423P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244517P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.

PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal

Sun Aug 15 07:51:05 2004

us-10-082-830-99.rng

cancers and also for testing and detection e.g. diagnosis.

Example 2: SEO ID NO 3654: 781pp + Sequence Listing; English.

The invention relates to novel genes (AAL34669-AAL37666) and proteins (AB03087-AB04109) associated with the musculoskeletal system useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct/sequences

Sequence 11027.BP; 2851 A; 2182 C; 2537 G; 3457 T; 0 U; 0 Other;

Query Match : 7.1%; Score 40; DB 4; Length 11027;

Query Match	7.1%;	Score 40;	22
Best Local Similarity	55.9%;	Pred. No. 2.7;	

Best Local Similarity	55.5%	Area: No: 2.7)	
Matches	76;	Conservative	0; Mismatches 60; Indels 0; Gaps 0;

417 GAAGAACAGGAGACAGAGACAAATCAAATGTCCTTACACAGAGACTGATTATATAGAT 476

5908 GAGGAACAAAAGCTTAAAGACAAATCAAAATGTCTCTAAATCAAT 5967

477 AATGGTACATTTCATATAAAACATGATATATTACTAATTAAACATACCACACACAAC 536

5968 TACCCCATTTCTCAGTAGAGTAAGATACAAGATACACAGCCTATTAAAGAAAGAA 6027

537 ACCTCGAGCCGCTAGT 552

6028 TGAAGGAGTTTCAAT 6043

Search completed: August 14, 2004, 19:13:46

search completed: August
Job time : 287.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 19:13:54 ; Search time 319.797 Seconds
(without alignments)
8638.053 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 ggcacattgtgtgtggat.....ggcgtagtctgagcttag 563

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	563	100.0	563	15	US-10-082-830-99
2	511.4	90.8	330973	13	Sequence 99, Appl
3	42.4	7.5	7374	10	Sequence 1498, Ap
4	42.4	7.5	7374	13	Sequence 10207, A
5	42	7.5	16766	15	Sequence 1006, Ap
6	42	7.5	744802	16	Sequence 2130, Ap
7	41.8	7.4	915	13	Sequence 1369, Ap
8	41.8	7.4	154817	16	Sequence 120056,
9	41.8	7.4	154817	16	Sequence 120056,
10	41.2	7.3	167343	9	Sequence 334, App
11	41.2	7.3	167343	9	Sequence 281, App
12	41	7.3	97415	17	Sequence 273, App
13	40.2	7.1	414	13	Sequence 287, App
14	40.2	7.1	414	16	Sequence 16076, A
					Sequence 16076, A

15	40.2	7.1	509	13	US-10-240-425-622	Sequence 622, App
16	40.2	7.1	145068	17	US-10-322-281-33	Sequence 33, Appl
17	40.2	7.1	155579	16	US-10-085-117-283	Sequence 283, App
18	40.2	7.1	173808	14	US-10-003-806-10	Sequence 10, Appl
19	40	7.1	714	13	US-10-027-632-114034	Sequence 114034,
20	40	7.1	714	13	US-10-027-632-114035	Sequence 114035,
21	40	7.1	714	13	US-10-027-632-114036	Sequence 114036,
22	40	7.1	714	16	US-10-027-632-114034	Sequence 114034,
23	40	7.1	714	16	US-10-027-632-114035	Sequence 114035,
24	40	7.1	714	16	US-10-027-632-114036	Sequence 114036,
25	40	7.1	11027	9	US-09-764-877-3654	Sequence 3654, Ap
26	40	7.1	11027	9	US-09-764-877-3655	Sequence 3655, Ap
27	40	7.1	11027	16	US-10-242-515-3654	Sequence 3654, Ap
28	40	7.1	11027	16	US-10-242-515-3655	Sequence 3655, Ap
29	40	7.1	104000	15	US-10-012-984-14	Sequence 14, Appl
30	40	7.1	104000	17	US-10-673-523-14	Sequence 14, Appl
31	39.8	7.1	878	13	US-10-027-632-31606	Sequence 31606, A
32	39.8	7.1	878	16	US-10-027-632-31606	Sequence 31606, A
33	39.6	7.0	911	13	US-09-854-867-341	Sequence 341, App
34	39.4	7.0	323	9	US-09-764-877-2658	Sequence 2658, Ap
35	39.4	7.0	323	9	US-09-764-877-2660	Sequence 2660, Ap
36	39.4	7.0	323	16	US-10-242-515-2658	Sequence 2658, Ap
37	39.4	7.0	323	16	US-10-242-515-2660	Sequence 2660, Ap
38	39.4	7.0	1771	15	US-10-181-590-18	Sequence 18, Appl
39	39.2	7.0	9725	15	US-10-311-455-1265	Sequence 1265, Ap
40	39.2	7.0	21537	15	US-10-311-455-1972	Sequence 1972, Ap
41	39	6.9	870	13	US-10-027-632-160076	Sequence 160076,
42	39	6.9	870	13	US-10-027-632-160077	Sequence 160077,
43	39	6.9	870	13	US-10-027-632-160078	Sequence 160078,
44	39	6.9	870	16	US-10-027-632-160076	Sequence 160076,
45	39	6.9	870	16	US-10-027-632-160077	Sequence 160077,

ALIGNMENTS

RESULT 1

US-10-082-830-99
; Sequence 99, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 99
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-082-830-99

Query Match 100.0%; Score 563; DB 15; Length 563;
Best Local Similarity 100.0%; Pred. No. 4.1e-147;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACATCTTGCTGGATGGAATTTGACATCTTTTCTTTTCTTCTGTTACAT 60
DB 1 GGTACATCTTGCTGGATGGAATTTGACATCTTTTCTTTTCTTCTGTTACAT 60
QY 61 ATCAAAATCTTAGATGTATTACTTCCAAAGCGGGTAAACTTATTCAAGATGACTGATC 120
DB 61 ATCAAAATCTTAGATGTATTACTTCCAAAGCGGGTAAACTTATTCAAGATGACTGATC 120

121	TCCTATTTTCCCTTAATTTTACTAGAGGTGAGGCTCCCAAGAACTACAGAAAATAGAAAGGAA	180
121	TCCTATTTTCCCTTAATTTTACTAGAGGTGAGGCTCCCAAGAACTACAGAAAATAGAAAGGAA	180
181	AGTCTCCATTGAGCCATCAACTGTGTGACACCTGGCATTTTAAGCATGAAGATAGGAGTAGGGCTTC	240
181	AGTCTCCATTGAGCCATCAACTGTGTGACACCTGGCATTTTAAGCATGAAGATAGGAGTAGGGCTTC	240
241	TATGTTAGGAGCTGGAGTAGGCAGCAGCATTTCCAGGAAAGGATCTCAGAGGTCAGAAAACAATA	300
241	TATGTTAGGAGCTGGAGTAGGCAGCAGCATTTCCAGGAAAGGATCTCAGAGGTCAGAAAACAATA	300
301	GATTATCAGTTAAATACTCTTGGACCAAGAGACCTTGAAATCCTCGCTCGGTGATGTA	360
301	GATTATCAGTTAAATACTCTTGGACCAAGAGACCTTGAAATCCTCGCTCGGTGATGTA	360
361	TACATATTTAATACACAATGAGAAAGCTCCTTTGAGGGTGAGTGAGAGTGATAGGGCTGAAG	420
361	TACATATTTAATACACAATGAGAAAGCTCCTTTGAGGGTGAGTGAGAGTGATAGGGCTGAAG	420
421	AACAGAGACAGAGAGACAATTCAAATGTCTCTTACACAGAGACTGATTATATAGATAATG	480
421	AACAGAGACAGAGAGACAATTCAAATGTCTCTTACACAGAGACTGATTATATAGATAATG	480
481	GTACATTTCAATATAAACATGATATATTTACTAATTTAAAAACATATCAACACACACCAACCT	540
481	GTACATTTCAATATAAACATGATATATTTACTAATTTAAAAACATATCAACACACACCAACCT	540
541	CGAGCGGCTAGTCTCGAGTCTAG	563
541	CGAGCGGCTAGTCTCGAGTCTAG	563

RESULT 2

```

RESULTS 2
US/10-087-192-1498
, Sequence 1498, Application US/10087192
, Publication No. US20020182586A1
, GENERAL INFORMATION:
, APPLICANT: Morris, David W.
, APPLICANT: Engelhard, Eric K.
, TITLE OF INVENTION: NOVEL COMPOSITIONS AND
, TITLE OF INVENTION: CANCER
, FILE REFERENCE: 529452000122
, CURRENT APPLICATION NUMBER: US/10/087,192
, CURRENT FILING DATE: 2002-03-01
, PRIOR APPLICATION NUMBER: US 09/747,377
, PRIOR FILING DATE: 2000-12-22
, PRIOR APPLICATION NUMBER: US 09/798,586
, PRIOR FILING DATE: 2001-03-02
, NUMBER OF SEQ ID NOS: 2059
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 1498
, LENGTH: 330973
, TYPE: DNA
, ORGANISM: Homo sapiens
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (1)..(330973)
, OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1498

```

	Query Match	90.8%;	Score 511.4;	DB 13;	Length 330973;
	Best Local Similarity	99.6%;	Pred. No. 4.2e-131;		
	Matches 523;	Conservative	0;	Mismatches 1;	Indels 1; Gaps 1;
QY	1	GGTACATCTTGGCGCTGGATGGAAATTTGACATACTTTTTATTATTTCTTTACCTGTTTACAT	60		
Db	147784	GGTACATCTTGGCGCTGGATGGAAATTTGACATACTTTTTATTATTTCTTTACCTGTTTACAT	147843		
QY	61	ATCAAACTCTTAGGATGTATTACTTCCAAAGGGGGTTAAACTATTCAAGATTGACTGAGTC	120		
Db	147844	ATCAAACTCTTAGGATGTATTACTTCCAAAGGGGGTTAAACTATTCAAGATTGACTGAGTC	147903		

Qy	121	TCTATTTTTCCTTAAATTTACTAGAACTGGAGCTCCAGAACTACACAGAAATAGAGGAA	180
Db	147904	TCTATTTTTCCTTAAATTTACTAGAACTGGAGCTCCAGAACTACACAGAAATAGAGGAA	147963
Qy	181	AGTCTCCATTGAGCCATGAAGCTGTGAGCACCTGGCATTTAAGCATGAAGAGTAGGCTTC	240
Db	147964	AGTCTCCATTGAGCCATGAAGCTGTGAGCACCTGGCATTTAAGCATGAAGAGTAGGCTTC	148023
Qy	241	TATGTAGGAGCTGGAGTAGGACAGCATTCGAGGAAGGATCTCAGAGGTCAGAAACAATA	300
Db	148024	TATGTAGGAGCTGGAGTAGGACAGCATTCGAGGAAGGATCTCAGAGGTCAGAAACAATA	148083
Qy	301	GATTATCAGTTAAATACTCTTGGACCAAGAAAGACCTTTGAAATCCTTGGCTCGTGATGTA	360
Db	148084	GATTATCAGTTAAATACTCTTGGACCAAGAAAGACCTTTGAAATCCTTGGCTCGTGATGTA	148143
Qy	361	TACATATTTAATACACAATGAGAGGCTCCTTGAGGGTGAGTGAGAGTGATAGGGCTGAAG	420
Db	148144	TACATATTTAATACACAATGAGAGGCTCCTTGAGGGTGAGTGAGAGTGATAGGGCTGAAG	148203
Qy	421	RACAGGACACAGACAGACAAATTCRAATCTCCTTACACAGAGAGCTGATTATATAGATAATG	480
Db	148204	RACAGGACACAGACAGACAAATTCRAATCTCCTTACACAGAGAGCTGATTATATAGATAATG	148263
Qy	481	GTACATTTCCATATAAACATGATATATTTACTAATTTAAACATACCA	525
Db	148264	GTACATTTCCATAT-AACATGATATATTTACTAATTTAAACATACCA	148307

RESULT 3

US-09-764-891-10207/c

; Sequence 10207, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10207

; LENGTH: 7374

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-891-10207

RESULT 3

```

US-09-764-891-10207/C
; Sequence 10207, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10207
; LENGTH: 7374
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-10207

```

```

Query Match      7.5%; Score 42.4; DB 10; Length 7374;
Best Local Similarity 57.6%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY      387 TCCTTGAGGGTGCAGTGCAGATGATAGGGCTGTGAAGAAACAGGAGACAGACAAATTCAAAT 446
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       1091 TACATGATGTTTCTTGAAGCATTTACATPATGACCAAAAAATGAAAATAATTCAAAT 1032

QY      447 GTCCTTACACAGAGACTGTTATATAGATAATGGTGACATTCATATAACATGATATATT 506
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       1031 GCCCATCAACTGCATGAATAGATAAATAAATGTGATATATCCATCAATGAAATGTTATT 972

QY      507 TACTAATTAAAA 518
          |||  |||  |||  |||
Db       971 CAGCAATAAAA 960

RESULT 4
US-10-205-428-1006/c
; Sequence 1006, Application US/10205428
; Publication NO. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAIL7C1

```

RESULT 4

```

RESULTS:
US-10-205-428-1006/c
; Sequence 1006, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PALL17C1

```



```
; CURRENT APPLICATION NUMBER: US/10/205.428
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1019
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1006
; LENGTH: 7374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-428-1006

Query Match      7.5%; Score 42.4; DB 15; Length 7374;
Best Local Similarity 57.6%; Pred. No. 1.1; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 56;

QY 387 TCCTTGAGGGTGAGTGAGTGATGAGGCTGAAGAACAGGAGACAGAGACAAATTCAAAT 446
Db 1091 TACATGAATGTTTGTTCAGCATTTATACATAATGACCAAAAAAATGAATATTCAT 1032

QY 447 GTCCTTACAGAGACATGATTTATATAGATAAAGGTACATTCATATAACATGATATTT 506
Db 1031 GTCCATCAACTGATGAATAGATAAATAAATGTCTATATCCATACATGAAATGTTATT 972

QY 507 TACTAATATAAAA 518
Db 971 CAGCAATAAAAA 960

RESULT 5
US-10-311-455-2130/c
; Sequence 2130, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2130
; LENGTH: 16766
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2130

Query Match      7.5%; Score 42; DB 15; Length 16766;
Best Local Similarity 60.5%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 45;

QY 422 ACAGGAGACAGAGACAAATTCAAATGCTCTTACACAGAGAGACTGATTATATAGATAATGG 481
Db 503 AAAAAAACTAAAAATAATTCAAATATCTATATAAATACTAATTATATAAATAATAA 444

QY 482 TACATTCATATAAACATGATATATTTACTAATTTAAAAACATACCAACACACACAA 535
Db 443 TATATCTATAAAATAAACTATATCCAACTATTTAAAAACAAAAATAAAAAA 390

RESULT 6
US-10-292-798-1369/c
; Sequence 1369, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-08-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1369
; LENGTH: 744802
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(744802)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(246)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25640)..(25677)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27078)..(27094)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141192)..(141769)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159571)..(159606)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174525)..(174575)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234891)..(235013)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (235514)..(235560)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (279677)..(279729)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (408660)..(409123)
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: (409204)..(409669)
FEATURE:
NAME/KEY: CDS
LOCATION: (428381)..(428396)
FEATURE:
NAME/KEY: CDS
LOCATION: (472204)..(472330)
FEATURE:
NAME/KEY: CDS
LOCATION: (714252)..(714355)
FEATURE:
NAME/KEY: CDS
LOCATION: (714447)..(714529)
FEATURE:
NAME/KEY: CDS
LOCATION: (739794)..(739891)
FEATURE:
NAME/KEY: CDS
LOCATION: (744484)..(744602)
FEATURE:
NAME/KEY: modified_base
LOCATION: (51812)..(51911)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (57122)..(57221)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (79368)..(79467)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (293951)..(294050)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (310089)..(310188)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (332935)..(332935)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (332992)..(332992)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (362002)..(362101)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (639781)..(639880)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1369
Query Match 7.5%; Score 42; DB 16; Length 744802;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
Qy 426 GAGCAGAGACAAATTCCTTACACAGAGACTGATATATAGATATGGTACA 485
Db 609737 GAAAGGAGCAATTCAGATGTCGTCATAGGGACTCTTTGGATAAATAATGCTGCT 609678
Qy 486 TTCATATAACATGATATATTTACTAATTA 515
Db 609677 TGCATCTATTCAGATATTTTAGATTA 609648
```

RESULT 7

```
US-10-027-632-120056/c
Sequence 120056, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 120056
LENGTH: 915
TYPE: DNA
ORGANISM: Human
US-10-027-632-120056
```

```
Query Match 7.4%; Score 41.8; DB 13; Length 915;
Best Local Similarity 63.4%; Pred. No. 0.49;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
Qy 418 AAGAACAGGACAGACAGACAAATTCCTTACACAGAGACTGATATATAGATA 477
Db 577 AATACCCCAACACAGTAAACAATCCAAAGACCCCATTCACAGGTGAAGAATAAATAATT 518
Qy 478 ATGGTACATTCATATAACATGATATATTTACTAATTAATAAAA 518
Db 517 ATGCTATATTCATACATAGATATCTATCTCACAATTAAGA 477
```

```
RESULT 8
US-10-027-632-120056/c
Sequence 120056, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 120056
```

; LENGTH: 915
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-120056

Query Match 7.4%; Score 41.8; DB 16; Length 915;
Best Local Similarity 63.4%; Pred. No. 0.49; 37; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGACAGACAGACAATTCCTTACACAGAGACTGATTATATAGATA 477
Db 577 AATAACCCCAACAGTAACAATCCAAAGACCCATTACAGGTGAAGAATAATAATT 518
QY 478 ATGCTACATTATATATAACATGATATATTTACTAATTAATAA 518
Db 517 ATGCTATATTCATACATAGATAATCTATCTCACAATTAAGA 477

RESULT 9

US-10-085-117-334/c
; Sequence 334, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:

; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 03/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 154817
; TYPE: DNA

; ORGANISM: Homo sapiens
; NAME/KEY: variation
; LOCATION: (1)...(154817)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-334

Query Match 7.4%; Score 41.8; DB 16; Length 154817;
Best Local Similarity 63.4%; Pred. No. 8.6;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGACAGACAGACAATTCCTTACACAGAGACTGATTATATAGATA 477
Db 88945 AACAGCCAAAAGTGGAAACAACTCAAATGTCCATCACTGATGAATGAATGATAAATT 88886
QY 478 ATGCTACATTATATATAACATGATATATTTACTAATTAATAA 518
Db 88885 GTGGTATATCTATATATGGAATATATTTAGCCATTAAAA 88845

RESULT 10

US-09-962-436-281
; Sequence 281, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:

; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 281
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-281

Query Match 7.3%; Score 41.2; DB 9; Length 167343;
Best Local Similarity 67.4%; Pred. No. 13;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 409 ATAGGGCTGAAGAACAGGAGACAGAGACAATTCAAATGTCTTACACAGAGACTGATT 468
Db 159637 ATTGTTCAAAATACAAAAGATTGGAAGAAGGCAATATCTTCAGTAGAAGACTGATG 159696
QY 469 ATATAGATAATGGTACATTCATATAA 494
Db 159697 AAATACATTGTGTACATACATACAA 159722

RESULT 11

US-09-964-824A-273
; Sequence 273, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:

; APPLICANT: Horriigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signat
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 273
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-273

Query Match 7.3%; Score 41.2; DB 9; Length 167343;
Best Local Similarity 67.4%; Pred. No. 13;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 409 ATAGGGCTGAAGAACAGGAGACAGAGACAATTCAAATGTCTTACACAGAGACTGATT 468
Db 159637 ATTGTTCAAAATACAAAAGATTGGAAGAAGGCAATATCTTCAGTAGAAGACTGATG 159696
QY 469 ATATAGATAATGGTACATTCATATAA 494
Db 159697 AAATACATTGTGTACATACATACAA 159722

RESULT 12

US-10-322-281-287
; Sequence 287, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:

; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 287
; LENGTH: 97415
; TYPE: DNA

```
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(97415)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-287

Query Match          7.1%; Score 41; DB 17; Length 97415;
Best Local Similarity 61.9%; Pred. No. 11;
Matches 65; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 416 TGAGACAGACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATTAATACGA 475
Db 48437 TTAATAGCCAGAGCTGGAAACACCCAGAAATTCCTCTCAACAGAAAGTGGATACAAAA 48496
Qy 476 TAATGGTACATTCATATAAACAATGATATATTCTAATTAATAACA 520
Db 48497 ATATGGTACATTTACACAATGGAGTACTCTTGTATTAACAACA 48541

RESULT 13
US-10-085-783A-16076/c
; Sequence 16076, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16076
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-16076

Query Match          7.1%; Score 40.2; DB 13; Length 414;
Best Local Similarity 59.0%; Pred. No. 0.88; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 239 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 180
Qy 469 ATATAGTAATGGTACATTCATATAAACAATGATATATTTACTAATTAACAATACCA 525
Db 179 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCGCTA 123

RESULT 14
US-10-242-535A-16076/c
; Sequence 16076, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
```

```
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16076
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-16076

Query Match          7.1%; Score 40.2; DB 16; Length 414;
Best Local Similarity 59.0%; Pred. No. 0.88; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 239 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 180
Qy 469 ATATAGTAATGGTACATTCATATAAACAATGATATATTTACTAATTAACAATACCA 525
Db 179 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCGCTA 123

RESULT 15
US-10-240-425-622
; Sequence 622, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 A142501.9
US-10-240-425-622

Query Match          7.1%; Score 40.2; DB 13; Length 509;
Best Local Similarity 59.0%; Pred. No. 0.98; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 263 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 322
Qy 469 ATATAGTAATGGTACATTCATATAAACAATGATATATTTACTAATTAACAATACCA 525
Db 323 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCGCTA 379

Search completed: August 15, 2004, 01:59:54
Job time : 322.797 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 11:00:56 ; Search time 54.0891 Seconds
(without alignments)
5776.350 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 ggtacatctgctgtgat.....gccgctagtctgagctag 563

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Databases : Issued Patents NA.*

- 1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/prodata/2/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	53.2	9.4	7218	1	US-08-232-463-14
C 2	39	6.9	81001	4	Sequence 14, Appl
C 3	37.4	6.6	408	4	Sequence 1, Appl
C 4	35.8	6.4	5535	4	Sequence 8793, Ap
C 5	35.6	6.3	112132	4	Sequence 18, Appl
C 6	35.6	6.3	112132	4	Sequence 3, Appl
C 7	35.4	6.3	3083	2	Sequence 36, Appl
C 8	35.4	6.3	3083	2	Sequence 36, Appl
C 9	35.4	6.3	3083	2	Sequence 36, Appl
C 10	35.4	6.3	3083	2	Sequence 36, Appl
C 11	35.4	6.3	3083	3	Sequence 36, Appl
C 12	35.4	6.3	3083	3	Sequence 36, Appl
C 13	35.4	6.3	3083	3	Sequence 36, Appl
C 14	35.4	6.3	3083	3	Sequence 36, Appl
C 15	35.4	6.3	3083	3	Sequence 36, Appl
C 16	35.4	6.3	3084	3	Sequence 11, Appl
C 17	35.4	6.3	3084	3	Sequence 11, Appl
C 18	35.4	6.3	3084	3	Sequence 6, Appl
C 19	35.4	6.3	3084	3	Sequence 11, Appl
C 20	35.4	6.3	3084	4	Sequence 6, Appl
C 21	35.4	6.3	3084	4	Sequence 11, Appl
C 22	35.4	6.3	3084	4	Sequence 11, Appl
C 23	35.4	6.3	3084	4	Sequence 11, Appl
C 24	35.4	6.3	29604	3	Sequence 13, Appl
C 25	35.4	6.3	29604	4	Sequence 207, App
C 26	34.8	6.2	6113	4	Sequence 14, Appl
C 27	34.6	6.1	11015	4	Sequence 55, Appl

C 28	34.2	6.1	832	4	US-09-621-976-2813	Sequence 2813, Ap
C 29	34.2	6.1	1313	4	US-09-149-476-112	Sequence 112, App
C 30	34.2	6.1	319608	4	US-09-539-333D-1	Sequence 1, Appl
C 31	34.2	6.1	319608	4	US-09-679-409-1	Sequence 1, Appl
C 32	34	6.0	3480	4	US-09-331-359-1	Sequence 1, Appl
C 33	34	6.0	19513	4	US-10-204-708-39	Sequence 39, Appl
C 34	34	6.0	83450	4	US-09-811-469-3	Sequence 3, Appl
C 35	33.8	6.0	3591	4	US-09-107-532A-290	Sequence 290, App
C 36	33.8	6.0	35100	1	US-08-306-891B-19	Sequence 19, Appl
C 37	33.8	6.0	35100	5	PCT-US93-06251-19	Sequence 17, Appl
C 38	33.8	6.0	176373	3	US-09-128-155-17	Sequence 66, Appl
C 39	33.6	6.0	4220	1	US-08-832-883-66	Sequence 66, Appl
C 40	33.6	6.0	4220	2	US-08-832-877-66	Sequence 352, App
C 41	33.6	6.0	9956	4	US-08-956-171E-352	Sequence 12, Appl
C 42	33.6	6.0	18073	3	US-09-078-294-12	Sequence 38, Appl
C 43	33.6	6.0	128779	4	US-09-497-855A-38	Sequence 11, Appl
C 44	33.6	6.0	392000	4	US-10-027-983-11	Sequence 1, Appl
C 45	33.4	5.9	4084	2	US-08-568-459A-1	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/POCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpc-Fls
US-08-232-463-14

Query Match

9.4%; Score 53.2; DB 1; Length 7218;


```
Db 2362 TTAAATC 2356
Query Match 6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

RESULT 5
US-09-741-150-3/c
; Sequence 3, Application US/09741150
; Patent No. 6436689
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000968
; CURRENT APPLICATION NUMBER: US/09/741.150
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 112132
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112132)
; OTHER INFORMATION: n = A,T,C or G
US-09-741-150-3

Query Match 6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 402 GAGAGTGATAGGGCTGAAGAACAGGACAGACAGCAATTCAAATGCTCTTACACAGAAG 461
DB 24899 GACAGCTTTATTTTAAATAGCCAAAACCTGAAACCAACCAAAATGCTATCAACAGATG 24840
QY 462 ACTGATTATATAGATAATGTTACATTCATATATACATGATATATTTACTAATTAACAAT 521
DB 24839 AATTGATAAACAATGATGGTATATCCATACACAGATACACTCCAGCAATAAAAAGA 24780
QY 522 ACCAACACAC 531
DB 24779 AACAAATTAC 24770

RESULT 7
US-08-480-994-36
; Sequence 36, Application US/08480994
; Patent No. 5834248
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/480,994
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-480-994-36

Db 2362 TTAAATC 2356
Query Match 6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

RESULT 6
US-10-160-187-3/c
; Sequence 3, Application US/10160187
; Patent No. 6620607
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000968DIV
; CURRENT APPLICATION NUMBER: US/10/160,187
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: 60/252,410
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: 09/741,150
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 112132
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112132)
; OTHER INFORMATION: n = A,T,C or G
US-10-160-187-3

QY 402 GAGAGTGATAGGGCTGAAGAACAGGACAGACAGCAATTCAAATGCTCTTACACAGAAG 461
DB 24899 GACAGCTTTATTTTAAATAGCCAAAACCTGAAACCAACCAAAATGCTATCAACAGATG 24840
QY 462 ACTGATTATATAGATAATGTTACATTCATATATACATGATATATTTACTAATTAACAAT 521
DB 24839 AATTGATAAACAATGATGGTATATCCATACACAGATACACTCCAGCAATAAAAAGA 24780
QY 522 ACCAACACAC 531
DB 24779 AACAAATTAC 24770
```


; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-599-654-36
Query Match 6.3%; Score 35.4; DB 2; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATA 477
Db 821 AACAGCCAAAGAGAGAGAGCAACCCAAATGTCCATAGCTGATTAATGGAAT 880
QY 478 ATGTACATTCATATAACATGATATATTTACTAATTAATAACATACCAACACACACA 537
Db 881 ATGTACATTCATATAACATGATATATTTACTAATTAATAACATACCAACACACA 537
QY 538 CTGAGCGGCTA 550
Db 941 AACGTGCTACAA 953
RESULT 11
US-08-944-868A-36
; Sequence 36, Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,868A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,654
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30

; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-599-654-36
Query Match 6.3%; Score 35.4; DB 2; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATA 477
Db 821 AACAGCCAAAGAGAGAGAGCAACCCAAATGTCCATAGCTGATTAATGGAAT 880
QY 478 ATGTACATTCATATAACATGATATATTTACTAATTAATAACATACCAACACACA 537
Db 881 ATGTACATTCATATAACATGATATATTTACTAATTAATAACATACCAACACACA 537
QY 538 CTGAGCGGCTA 550
Db 941 AACGTGCTACAA 953
RESULT 10
US-08-485-573-36
; Sequence 36, Application US/08485573
; Patent No. 5968770
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,573
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; US-08-944-868A-36

Query Match
Best Local Similarity 6.3%; Score 35.4; DB 3; Length 3083;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACACATTCAAATGCTTACACAGAGACTGATTATATAGATA 477
Db 821 AACACCCAAAGAGAGAGACCAACCCAAATGCTTACATAGCTGATAAATGATAAATGAAAT 880

QY 478 ATGGTACATTCATATAAACAATGATATATTTACTAATTAATAACATACCAACACACACA 537
Db 881 ATGGTACGTCGAGAGATGGAATATCATTCACCCATGAAAAGAACGAAATCCAGCACCA 940

QY 538 CCTCAGCGCGCTA 550
Db 941 AAACGTGCTACAA 953

RESULT 12
US-08-944-423A-36
; Sequence 36, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944.423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; FEATURE:

;
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; US-08-944-423A-36

Query Match
Best Local Similarity 6.3%; Score 35.4; DB 3; Length 3083;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACACATTCAAATGCTTACACAGAGACTGATTATATAGATA 477
Db 821 AACACCCAAAGAGAGAGACCAACCCAAATGCTTACATAGCTGATAAATGATAAATGAAAT 880

QY 478 ATGGTACATTCATATAAACAATGATATATTTACTAATTAATAACATACCAACACACACA 537
Db 881 ATGGTACGTCGAGAGATGGAATATCATTCACCCATGAAAAGAACGAAATCCAGCACCA 940

QY 538 CCTCAGCGCGCTA 550
Db 941 AAACGTGCTACAA 953

RESULT 13
US-08-925-743-36
; Sequence 36, Application US/08925743
; Patent No. 6054558
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,573
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
```



```
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
;
US-08-925-767-36

Query Match      6.3%; Score 35.4; DB 3; Length 3083;
Best Local Similarity 54.1%; Pred.No.1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 418 AAGACAGGAGACAGACGAATTCAATGTCCTTACACAGAGACTGATTATAGATA 477
Db 821 AACAGCCAAAGAGAGGAGCAACCCAAATGTCATTAGCTGATAAATGGATAATGAAT 880
Qy 478 ATGGTACATTTCATATAAACAATGATATATTACTAATTAAACATACCACACACACAACA 537
Db 881 ATGGTACGTCGGAAGATGGATATCATCCCATGAAAAGAACGAGTCCAGCACCA 940
Qy 538 CCTCGAGCCGCTA 550
Db 941 AAACGTGCTACAA 953
```

Search completed: August 14, 2004, 20:48:11
Job time : 56.0891 secs

This Page Blank (uspto)

QY 1 FFFFLRQSFTLSQAGVAWHDLGSLHPLPG 31
 ||||| : : : : : ||||| : : ||
 DB 311 FFFFEFESHVTOAGVQWRDLGSLQAPPG 341

RESULT 3

F40201
 artifact-warning sequence (translated ALU class F) - human
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 A:Accession: F40201
 R:Claverie, J.M.
 personal communication, 1992
 A:Reference number: A40201
 A:Accession: F40201
 A:Molecule type: DNA
 A:Residues: 1-673 <CLA>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891; PMID:1572661
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 33.7%; Score 83.5; DB 4; Length 673;
 Best Local Similarity 43.6%; Pred. No. 0.01; Mismatches 21; Indels 7; Gaps 1;
 Matches 24; Conservative 3; Mismatches 21; Indels 7; Gaps 1;

QY 1 FFFFLRQSFTL-----SQAGVAWHDLGSLHPLPGSSDRASASQSRITGV 48
 ||||| : : : : : ||||| : : ||
 DB 344 FFFFCFVFVFETGPCSDTPAGVQWQLAHCSNLNLGSSDSPASVSRVAGITGM 398

RESULT 4

A40201
 artifact-warning sequence (translated ALU class A) - human
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 A:Accession: A40201
 R:Claverie, J.M.
 personal communication, 1992
 A:Reference number: A40201
 A:Accession: A40201
 A:Molecule type: DNA
 A:Residues: 1-627 <CLA>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891; PMID:1572661
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 33.5%; Score 83; DB 4; Length 627;
 Best Local Similarity 51.6%; Pred. No. 0.011; Mismatches 10; Indels 0; Gaps 0;
 Matches 16; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAWHDLGSLHPLPG 31
 ||||| : : : : : ||||| : : ||
 DB 321 FLFFSETSRVAKAGVQWRDLGSLQAPPG 351

RESULT 5

T02670
 probable thromboxane A2 receptor isoform beta - human
 C:Species: Homo sapiens (man)
 C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 15-Sep-2000
 C:Accession: T02670

R:Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; G
 J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; At
 P.; Quan, G.; Krommiller, B.; Arellano, A.; Montgomery, M.; Ow, D.; Nolan, M.
 submitted to the EMBL Data Library, June 1998

A:Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.
 A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine
 A:Reference number: Z14696
 A:Accession: T02670
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <LAM>
 A:Cross-references: EMBL:AC005175; NID:G3253116; PID:G3253117
 C:Superfamily: prostaglandin E receptor EPI

Query Match 32.1%; Score 79.5; DB 2; Length 407;
 Best Local Similarity 49.0%; Pred. No. 0.019; 6; Indels 13; Gaps 3;
 Matches 24; Conservative 6; Mismatches 13; Indels 13; Gaps 3;

QY 7 RQSFTLSQAGVAWHDL---GSL-----HPPLFGSSDRASASQSRITGV 48
 ||||| : : : : : ||||| : : ||
 DB 328 RSLTL-----WPSLEYSGTISAHCNLRPLPGSSDRASASRAAGITGV 370

RESULT 6

A56194
 thromboxane A-2 receptor, endothelial - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 07-May-1999
 C:Accession: A56194
 R:Raychowdhury, M.K.; Yukawa, M.; Collins, L.J.; McGrail, S.H.; Kent, K.C.; Ware, J.A.
 J. Biol. Chem. 270, 7011, 1995
 A:Reference number: A56194; MUID:95204505; PMID:7896853
 A:Accession: A56194
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-73 <RAY>
 A:Cross-references: GB:U11271
 C:Superfamily: prostaglandin E receptor EPI

Query Match 31.9%; Score 79; DB 2; Length 79;
 Best Local Similarity 85.0%; Pred. No. 0.0038; Mismatches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 29 LPGSSDRASASQSRITGV 48
 ||||| : : : : : ||||| : : ||
 DB 23 LPGSSDRASASRAAGITGV 42

RESULT 7

E40201
 artifact-warning sequence (translated ALU class E) - human
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 A:Accession: E40201
 R:Claverie, J.M.
 personal communication, 1992
 A:Reference number: A40201
 A:Accession: E40201
 A:Molecule type: DNA
 A:Residues: 1-597 <CLA>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891; PMID:1572661
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 30.0%; Score 74.5; DB 4; Length 597;
 Best Local Similarity 71.4%; Pred. No. 0.13; 2; Indels 1; Gaps 1;
 Matches 15; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

serotonin receptor 6 - human
N:Alternate names: 5-hydroxytryptamine receptor 6 (5-HTR6)
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence revision 05-Sep-1997 #text change

C;Accession: JG5520
 R;Kohan, R.; Mccall, M.A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J.E.; Meltzer, J. Neurochem. 66, 47-56, 1996
 A;Title: Cloning, characterization, and chromosomal localization of a human 5-HT 6 serotonergic receptor
 A;Reference number: JG5520; MUID:96102917; PMID:8522988
 A;Accession: JG5520
 A;Molecule type: mRNA
 A;Residues: 1-440 <KOR>
 A;Cross-references: GB:L41147; NID:g1162923; PIDN:AAA92622.1; PID:g1162924
 A;Experimental source: brain
 C;Comment: This protein shows high affinity for several therapeutically important antidiabetic drugs
 C;Genetics:
 A;Gene: HTR6
 A;Map position: 1p35-36
 A;Introns: 238/3: 291/3
 C;Superfamily: vertebrate rhodopsin
 C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transmembrane protein
 F;28-51/Domain: transmembrane #status predicted <TM1>
 F;65-88/Domain: transmembrane #status predicted <TM2>
 F;100-123/Domain: transmembrane #status predicted <TM3>
 F;143-166/Domain: transmembrane #status predicted <TM4>
 F;185-208/Domain: transmembrane #status predicted <TM5>
 F;266-289/Domain: transmembrane #status predicted <TM6>
 F;298-321/Domain: transmembrane #status predicted <TM7>
 F;10/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.8%; Score 59; DB 2; Length 440;
 Best Local Similarity 45.5%; Pred. No. 9.1;
 Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 17 VAWHDLGSLHPLPGSSDSRAS 38
 DB 164 LGWELGHARPEVPGQCLLAS 185

RESULT 13
 I78885
 serine/threonine-specific protein kinase (EC 2.7.1.-) STK2 - human
 C;Species: Homo sapiens (man)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999
 C;Accession: I78885
 R;Lavedakou, E.N.; He, M.; Baptist, E.W.; Craven, R.J.; Cance, W.G.; Welcsh, P.L.; Simmonds, J.; et al. J. Biol. Chem. 274, 1977-1988, 1994
 A;Title: Two novel human serine/threonine kinases with homologies to the cell cycle regulatory protein p34^{cdc2}
 A;Reference number: I58396; MUID:94268838; PMID:8208544
 A;Accession: I78885
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-841 <RES>
 A;Cross-references: GB:L20321; NID:g348244; PIDN:AAA36658.1; PID:g348245
 C;Genetics:
 A;Gene: GDB:STK2
 A;Cross-references: GDB:374125
 A;Map position: 3p21.1-3p21.1
 C;Superfamily: human serine/threonine-specific protein kinase homology
 C;Keywords: phosphotransferase
 F;4-261/Domain: protein kinase homology <KIN>

Query Match 23.8%; Score 59; DB 1; Length 841;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 29 LPGSSDSRASASQASITGV 48
 DB 478 LLGSSDSPASASRVAGITGV 497

RESULT 14
 AH2704
 methyltransferase Atu1041 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C;Species: Agrobacterium tumefaciens
 C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C;Accession: AH2704

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; et al. Science 294, 2317-2323, 2001
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A;Reference number: AB2577; MUID:21608550; PMID:11743193
 A;Accession: AH2704
 A;Status: Preliminary
 A;Molecule type: DNA
 A;Residues: 1-244 <KUR>
 A;Cross-references: GB:AE008688; PIDN:AAL42054.1; PID:g17739432; GSPDB:GN00186
 A;Experimental source: strain C58 (Dupont)
 C;Genetics:
 A;Gene: Atu1041
 A;Map position: circular chromosome

Query Match 23.4%; Score 58; DB 2; Length 244;
 Best Local Similarity 37.1%; Pred. No. 6.5;
 Matches 23; Conservative 5; Mismatches 16; Indels 18; Gaps 4;

QY 4 FTLROSFT-----LSQAGVAWHDLGSLHPLPGSS-----DSRASASQ-SARIT 46
 DB 11 FFERYSAMPRTSEGLRQAG-EWHELRLAMPDLKGRSFLDLGCGFGWHCRVAEQGAARIV 69

QY 47 GV 48
 DB 70 GV 71

RESULT 15
 H97486
 biotin synthesis protein bioC VC1114 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C;Species: Agrobacterium tumefaciens
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C;Accession: H97486
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Williams, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; et al. Science 294, 2323-2328, 2001
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A;Reference number: A97359; MUID:21608551; PMID:11743194
 A;Accession: H97486
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-244 <KUR>
 A;Cross-references: GB:AE007869; PIDN:AAK86849.1; PID:g15156063; GSPDB:GN00169
 C;Genetics:
 A;Gene: AGR_C 1920
 A;Map position: circular chromosome

Query Match 23.4%; Score 58; DB 2; Length 244;
 Best Local Similarity 37.1%; Pred. No. 6.5;
 Matches 23; Conservative 5; Mismatches 16; Indels 18; Gaps 4;

QY 4 FTLROSFT-----LSQAGVAWHDLGSLHPLPGSS-----DSRASASQ-SARIT 46
 DB 11 FFERYSAMPRTSEGLRQAG-EWHELRLAMPDLKGRSFLDLGCGFGWHCRVAEQGAARIV 69

QY 47 GV 48
 DB 70 GV 71

Search completed: August 10, 2004, 19:15:42
 Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004, Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:11:21 ; Search time 13 Seconds
(without alignments)
192.259 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSTLSAGVAWH.....LPQSSDSRASQSRITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	43.5	591	ALU8_HUMAN	P39135 homo sapien
2	103.5	41.7	593	ALU7_HUMAN	P39134 homo sapien
3	97	39.1	603	ALU4_HUMAN	P39131 homo sapien
4	96.5	38.9	593	ALU6_HUMAN	P39133 homo sapien
5	83	33.5	587	ALU2_HUMAN	P39139 homo sapien
6	83	33.5	587	ALU3_HUMAN	P39130 homo sapien
7	77	31.0	585	ALU5_HUMAN	P39132 homo sapien
8	76.5	30.8	591	ALU1_HUMAN	P39138 homo sapien
9	73	29.4	438	SCBE_HUMAN	O43556 homo sapien
10	67	27.0	1156	HTS_DROME	Q02645 drosophila
11	65.5	26.4	741	RED1_HUMAN	P78563 homo sapien
12	65	26.2	211	BCS4_HUMAN	O8tdm0 homo sapien
13	63	25.4	629	2195_HUMAN	O14628 homo sapien
14	62	25.0	833	MAK1_HUMAN	Q32918 homo sapien
15	61	24.6	726	VNCS_PAVBO	P07236 bovine parv
16	59	23.8	440	SH6_HUMAN	P50406 homo sapien
17	59	23.8	841	NEK4_HUMAN	P51957 homo sapien
18	58	23.4	407	ADRM_HUMAN	Q16186 homo sapien
19	57	23.0	811	Y104_SYNY3	P54371 synecocyst
20	56	22.6	902	GLR4_HUMAN	P48058 homo sapien
21	55	22.2	2326	CCAB_DISOM	P56698 discopyse o
22	55	22.2	2774	NAPA_RAT	P34926 rattus norv
23	54	21.8	2333	PGCA_CANPA	Q28343 canis fami
24	54	21.8	2805	NAPA_HUMAN	P78559 homo sapien
25	53.5	21.6	167	FTMG_ECOLI	P08190 escherichia
26	53.5	21.6	271	S24D_ANOGA	O17004 anopheles g
27	53.5	21.6	847	NAT3_HUMAN	P43243 homo sapien
28	53	21.4	176	NLI2_PABJU	Q04404 parietaria
29	53	21.4	653	APPI_MOUSE	Q03157 mus musculu
30	53	21.4	977	REMF_HUMAN	Q96t37 homo sapien
31	52.5	21.2	348	OMBI_NEIGO	P18195 neisseria g
32	52.5	21.2	350	MRP_DEIRA	O9vm9 deinococcus
33	52.5	21.2	402	B4G1_BOVIN	P08037 b beta-1,4-

34 52 21.0 407 1 ADRM_MOUSE
35 52 21.0 407 1 ADRM_RAT
36 52 21.0 476 1 EGR2_HUMAN
37 52 21.0 663 1 GLI3_CHICK
38 52 21.0 710 1 IRL1_MOUSE
39 52 21.0 769 1 TFR1_FELCA
40 52 21.0 1011 1 M3K6_HUMAN
41 52 21.0 1183 1 DRPL_RAT
42 52 21.0 1235 1 CYA4_TRYBB
43 51.5 20.8 382 1 Y894_HAEIN
44 51.5 20.8 487 1 WRK1_ARATH
45 51 20.6 226 1 IBP6_RAT

Q9jkw1 mus musculu
Q9jmb5 rattus norv
P11161 homo sapien
P55879 gallus gall
Q62406 mus musculu
Q9myz3 felis silve
P54382 homo sapien
Q26721 trypanosoma
Q57500 haemophilus
Q9s137 arabidopsis
P35572 rattus norv

ALIGNMENTS

RESULT 1
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39135;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Miliusavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX', IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons,

consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

-!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14574; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 195 FRAME-2.

FT DOMAIN 199 294 FRAME-3.

FT DOMAIN 298 393 FRAME-4.

FT DOMAIN 397 492 FRAME-5.

FT DOMAIN 496 591 FRAME-6.

FT SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Query Match 43.5%; Score 108; DB 1; Length 591;

Best Local Similarity 32.2%; Pred. No. 6.1e-06;

Matches 28; Conservative 6; Mismatches 11; Indels 42; Gaps 1;

31

QY 4 FFLRQSFTLSQAGVAVHDLGSLHPPPLG----- 31

Db 496 FFEESRSVAQGVQWRDLGSLQPPFPGRFSCSLPSSWDYRRAPPNPANFCIFSRDG 555

QY 32 -----SSDSRASASQARITGV 48

Db 556 VSPCPWGSRTPTDLRXSSQSAGITGV 582

RESULT 2

ALU7_HUMAN STANDARD; PRT; 593 AA.

AC P39194;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alu subfamily SQ sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RP MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M.; Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752 (1994).

RN [2]

RP CONCEPT.

RP MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL Genomics 12:838-841 (1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RP MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

J. Mol. Evol. 27:194-202 (1988).

[4]

RN ALU FAMILIES CLASSIFICATION.

RP MEDLINE=91178915; PubMed=1706781;

RX Jurka J., Milosavljevic A.;

RA "Reconstruction and analysis of human Alu genes.";

RT J. Mol. Evol. 32:105-121 (1991).

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.

CC -!- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

CC -!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14573; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 97 FRAME-1.

FT DOMAIN 101 196 FRAME-2.

FT DOMAIN 200 295 FRAME-3.

FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.

FT DOMAIN 498 593 FRAME-6.

FT SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Query Match 41.7%; Score 103.5; DB 1; Length 593;

Best Local Similarity 72.4%; Pred. No. 2.2e-05;

Matches 21; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

31

QY 4 FFLRQSFTLSQAGVAVHDLGSLHPPPLG 31

Db 299 FFEESRSFALVAQGVQWRDLGSLQPPFPFG 327

RESULT 3

ALU4_HUMAN STANDARD; PRT; 603 AA.

ID ALU4_HUMAN

AC P39191;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alu subfamily SB2 sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

RT "Alu alert.";
 RL Nature 371:752-752 (1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL Genomics 12:838-841 (1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202 (1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121 (1991).
 CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -!- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -!- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14569; -; NOT_ANNOTATED_CDS.
 DR Hypothetical protein.
 KW DOMAIN 1 96 FRAME-1.
 FT DOMAIN 100 194 FRAME-2.
 FT DOMAIN 198 292 FRAME-3.
 FT DOMAIN 296 391 FRAME-4.
 FT DOMAIN 395 489 FRAME-5.
 FT DOMAIN 493 587 FRAME-6.
 SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;
 Query Match 33.58; Score 83; DB 1; Length 587;
 Best Local Similarity 57.18; Pred. No. 0.0082;
 Matches 16; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 FFILRQSTLSQAGVAMHDIGSLHPLPG 31
 || : : : ||| ||| ||| |||
 Db 493 FFETESRVAQAGVQWRDLGSLQAPPFG 520
 RESULT 7
 ID -ALUS_HUMAN STANDARD; PRT; 585 AA.
 AC P39192;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alu subfamily SC sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert.";
 RL Nature 371:752-752 (1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL Genomics 12:838-841 (1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202 (1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121 (1991).
 CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -!- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -!- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL; U14571; -; NOT ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 95
CC FT DOMAIN 99 193
CC FT DOMAIN 197 291
CC FT DOMAIN 295 389
CC FT DOMAIN 393 487
CC FT DOMAIN 491 585
CC FT SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;
SQ
Query Match 31.0%; Score 77; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 0.046;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 14 QAGVAVHDLGSLHPLPG 31
DB 403 QAGVQWRDLGSLQPPPG 420
RESULT 8
ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178915; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON 'XXX'. IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
```

```
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL; U14567; -; NOT ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 96
CC FT DOMAIN 100 195
CC FT DOMAIN 199 294
CC FT DOMAIN 298 393
CC FT DOMAIN 397 492
CC FT DOMAIN 496 591
CC FT SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;
SQ
Query Match 30.8%; Score 76.5; DB 1; Length 591;
Best Local Similarity 47.9%; Pred. No. 0.054;
Matches 23; Conservative 4; Mismatches 10; Indels 11; Gaps 2;
QY 2 FFFFLRQSF-TLSQAGVAVHDLGSLHPLPGSSDSRASASQASRITGV 48
DB 351 FVFFVETGFHYVQAQGL-----ELLGSSDPPASASQASGITGV 388
RESULT 9
SGCE_HUMAN STANDARD; PRT; 438 AA.
ID SGCE_HUMAN
AC O43556;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Epsilon-sarcoglycan precursor (Epsilon-SG).
GN SGCE OR ESG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Nigro V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA McNally E.M., Ly C.T., Kunkel L.M.;
RT "Human epsilon-sarcoglycan is highly related to alpha-sarcoglycan
RT (adhalin), the limb girdle muscular dystrophy 2D gene.";
RL FEBS Lett. 422:27-32(1998).
```



```

CC CC -!- SIMILARITY: Contains 2 DBM (double-stranded RNA-binding) domains.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U82120; AAB61686.1; -
DR EMBL; U82121; AAB61687.1; -
DR EMBL; X99227; CAA67611.1; -
DR EMBL; X99383; CAA67762.1; -
DR EMBL; U76420; AAC51240.1; -
DR EMBL; U76421; AAC51241.1; -
DR EMBL; U76422; AAC51242.1; -
DR EMBL; AF001042; AAB58300.1; -
DR EMBL; AL163301; CAB90493.1; -
DR HSSP; Q91836; 1D12.
DR Genew; HGNC:226; ADARB1.
DR GK; P78563; -
DR MIM; 601218; -
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0004000; P:adenosine deaminase activity; TAS.
DR GO; GO:0003725; P:double-stranded RNA binding; TAS.
DR GO; GO:0007417; P:central nervous system development; TAS.
DR GO; GO:0006396; P:RNA processing; TAS.
DR InterPro; IPR002466; A:deamin.
DR InterPro; IPR001159; DS:RBD.
DR Pfam; PF02137; A:deamin; 1.
DR Pfam; PF00035; dsrm; 2.
DR SMART; SM00552; ADEAMC; 1.
DR SMART; SM00358; DSRM; 2.
DR PROSITE; PS50141; A:DEAMIN EDITASE; 1.
DR PROSITE; PS50137; DS_RBD; 2.
KW mRNA processing; Hydroxylase; Zinc; RNA-binding; Repeat;
KW Alternative splicing.
FT DOMAIN 76 146
FT DOMAIN 230 300
FT METAL 394 394
FT ACT_SITE 396 396
FT METAL 451 451
FT METAL 556 556
FT VARSPPLIC 466 505
FT VARSPPLIC 713 741
FT FTID=VSP_000865.
FT FTID=VSP_000866.
FT FTID=VSP_000867.
FT G -> A (IN REF. 4).
FT R -> E (IN REF. 4).
FT V -> L (IN REF. 4).
SQ SEQUENCE 741 AA; 80763 MW; 02B583414DD59C20 CRC64;
Query Match 26.4%; Score 65.5; DB 1; Length 741;
Best Local Similarity 45.2%; Pred. No. 1.7;
Matches 19; Conservative 3; Mismatches 19; Indels 1; Gaps 1;
Qy 6 LRQFTLSQAGVAMHDLGLSLPLFG-SSDSRASASQARIT 46
Db 464 LEGRSVYQAGVQVQCNHGLQRPRLPGLLSPSTFTFGAGTT 505
RESULT 12
ID BC54 HUMAN STANDARD; PRT; 211 AA.
AC Q8TDM0; Q8IX17; Q8NEZ6; Q8TDL9; Q9NX13; Q9Y511;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Breast carcinoma amplified sequence 4.
GN BCAS4.
OS Homo sapiens (Human).

```

```

CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC CC [1]
RN SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND CHROMOSOMAL
RN TRANSLOCATION WITH BCAS3.
RN TISSUE=Liver;
RX MEDLINE=22265331; PubMed=12378525;
RA Baerlund M., Monni O., Weaver J.D., Kauraniemi P., Sauter G.,
RA Heiskanen M., Monni O.-P., Kallioniemi A.;
RT "Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo
RT amplification, overexpression, and fusion in breast cancer.";
EL Genes Chromosomes Cancer 35:311-317(2002).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Guillam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharshaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Wittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RN SEQUENCE OF 14-211 FROM N.A. (ISOFORM 1).
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obaishi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RN SEQUENCE OF 27-211 FROM N.A. (ISOFORM 2).
RA Bauer M.;
RT "Cloning and sequencing of a new BCAS4 isoform (BCAS4 isoform 1).";
RL Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Name=1;
CC CC IsoId=Q8TDM0-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId=Q8TDM0-2; Sequence=VSP_007854, VSP_007855;
CC CC -!- TISSUE SPECIFICITY: Brain, thymus, spleen, kidney and placenta.
CC CC Overexpressed in most breast cancer cell lines.
CC CC -!- DISEASE: May be involved in breast cancer through a
CC CC t(17;20)(q23;q13) chromosomal translocation that involves BCAS3
CC CC and BCAS4.
CC CC -!- SIMILARITY: Belongs to the cappuccino family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way

```



```

CC -!- FUNCTION: May play a role in hematopoietic lineage decisions and
CC growth regulation. (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- COFACTOR: Magnesium.
CC -!- TISSUE SPECIFICITY: Expressed primarily in hematopoietic organs,
CC including bone marrow, spleen and thymus. Also expressed at very
CC low levels in lung, kidney, mammary glands and small
CC intestine.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC STE20 subfamily.
CC -!- SIMILARITY: Contains 1 CNH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U66464; AAB97983.1; -
CC HSPSP; O63450; IAO6
CC Gnew; HGNC:6863; MAP4K1.
CC MIM; 601983; -
CC GO; GO:0005524; F:ATP binding; IDA.
CC GO; GO:0004574; F:protein serine/threonine kinase activity; IDA.
CC GO; GO:0007257; P:activation of JUNK; TAS.
CC GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
CC GO; GO:0007243; P:protein kinase cascade; IDA.
CC GO; GO:0045610; P:regulation of hemocyte differentiation; ISS.
CC GO; GO:0006950; P:response to stress; IDA.
CC InterPro; IPR001180; Citron.
CC InterPro; IPR000719; Prot kinase.
CC InterPro; IPR008271; Ser thr pkin AS.
CC InterPro; IPR002290; Ser thr pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00780; CNH; 1.
CC PRINTS; PR00069; pkinase; 1.
CC PRODOM; PD000001; Prot kinase; 1.
CC SMART; SM00036; CNH; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN KINASE_ST; FALSE NEG.
CC ATP-binding; Transferase; Kinase; Serine/threonine-protein kinase.
CC DOMAIN 17 274 PROTEIN KINASE.
CC DOMAIN 501 807 CNH.
CC NP_BIND 23 31 ATP (BY SIMILARITY).
CC BINDING 46 46 ATP (BY SIMILARITY).
CC ACT_SITE 137 137 BY SIMILARITY.
CC SEQUENCE 833 AA; 91296 MW; 3C98CF01BE42E151 CRC64;
Query Match 25.0%; Score 62; DB 1; Length 833;
Best Local Similarity 53.3%; Pred. No. 5.3;
Matches 16; Conservative 5; Mismatches 5; Indels 4; Gaps 1;
Qy 23 GSHLHP-----LPGSSDSRASASQARITGV 48
Db 804 GTISPHCNLLLPGSSNSPASASRVAGITGL 833
RESULT 15
VNCS_PAVEO
ID_VNCS_PAVBO STANDARD; PRT; 726 AA.
AC P07296;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Noncapsid protein NS-1 (Nonstructural protein NS1).
GN NS1.
OS Bovine parvovirus (BPV).

```

```

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10784;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87061184; PubMed=3783814;
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,
RA Bates R.C.;
RT "Complete nucleotide sequence and genome organization of bovine
RT parvovirus.";
RL J. Virol. 60:1085-1097(1986).
CC -!- SIMILARITY: Belongs to the parvoviruses noncapsid protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14363; AAB59845.1; -
CC PIR; C26104; UYPVSI.
CC InterPro; IPR001257; Parvo NS1.
CC Pfam; PF01057; Parvo NS1; 1.
CC Nonstructural protein; Noncapsid protein; DNA replication;
CC ATP-binding. 310 317 ATP (POTENTIAL)..
CC NP_BIND 726 AA; 81189 MW; 42659A611A3AF038 CRC64;
SQ SEQUENCE 726 AA; 81189 MW; 42659A611A3AF038 CRC64;
Query Match 24.6%; Score 61; DB 1; Length 726;
Best Local Similarity 30.0%; Pred. No. 6;
Matches 12; Conservative 6; Mismatches 22; Indels 0; Gaps 0;
Qy 8 QSFTLSQAGVAWHDGLSLHPLPGSSDSRASASQARITG 47
Db 486 QNWTFSENGVCVHCQGFQPTPESDTDSGDPPDPAVAG 525
Search completed: August 10, 2004, 19:14:28
Job time : 14 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: August 10, 2004, 19:11:46 ; Search time 35 seconds
(without alignments)
432.711 Million cell updates/sec

Title: US-10-082-830-224
Perfect score: 248
Sequence: 1 FFFFLQSFLLSQAGVAWH.....LPSSDSRASASQARITGV 48

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	58.1	100	4	Q7826 homo sapien
2	135	54.4	109	4	Q8N181 homo sapien
3	131	52.8	152	4	Q9NX85 homo sapien
4	124	50.0	208	4	Q9NW14 homo sapien
5	122.5	49.4	122	6	Q9BGW3 macaca fasc
6	122	49.2	121	4	Q8N8C8 homo sapien
7	121	48.8	162	4	Q8NAD5 homo sapien
8	118	47.6	158	4	Q8NAL9 homo sapien
9	117	47.2	118	4	Q9H3B7 homo sapien
10	108.5	43.8	138	4	Q8N874 homo sapien
11	106	42.7	157	4	Q8N845 homo sapien
12	105.5	42.5	535	4	Q96EB1 homo sapien
13	105	42.3	162	4	Q8H5D5 homo sapien
14	103	41.5	61	4	Q9UI48 homo sapien
15	102.5	41.3	135	6	Q95KE1 macaca fasc
16	102	41.1	133	4	Q96JRS homo sapien

17	101.5	40.9	375	4	O60448	O60448 homo sapien
18	101	40.7	93	6	O8HXK0	O8HXK0 macaca fasc
19	101	40.7	826	4	O00420	O00420 homo sapien
20	100.5	40.5	180	4	Q8N7M7	Q8N7M7 homo sapien
21	100	40.3	111	4	Q8NSM6	Q8NSM6 homo sapien
22	100	40.3	163	4	Q9GVM0	Q9GVM0 homo sapien
23	100	40.3	175	4	Q8N980	Q8N980 homo sapien
24	98.5	39.7	151	4	Q8N287	Q8N287 homo sapien
25	97.5	39.3	244	4	Q8H700	Q8H700 homo sapien
26	94.5	38.1	171	4	Q8H728	Q8H728 homo sapien
27	94	37.9	180	4	Q8NAQ6	Q8NAQ6 homo sapien
28	92.5	37.3	86	4	Q8NSB5	Q8NSB5 homo sapien
29	92.5	37.3	156	4	Q8NBH4	Q8NBH4 homo sapien
30	92	37.1	170	4	Q8NAI3	Q8NAI3 homo sapien
31	90.5	36.5	1150	4	Q8HBC4	Q8HBC4 homo sapien
32	89.5	36.1	239	4	Q8NX17	Q8NX17 homo sapien
33	89	35.9	140	4	Q8N891	Q8N891 homo sapien
34	89	35.9	156	4	Q8N266	Q8N266 homo sapien
35	88	35.5	179	4	Q8NB20	Q8NB20 homo sapien
36	85.5	34.5	102	4	Q8N4K1	Q8N4K1 homo sapien
37	85.5	34.5	127	4	Q8N8F4	Q8N8F4 homo sapien
38	83.5	33.7	84	4	Q9UHT1	Q9UHT1 homo sapien
39	83	33.5	121	4	Q9HA45	Q9HA45 homo sapien
40	83	33.5	151	4	Q9HA67	Q9HA67 homo sapien
41	82	33.1	126	4	Q9UHT2	Q9UHT2 homo sapien
42	80	32.3	27	4	Q96HX1	Q96HX1 homo sapien
43	79.5	32.1	407	4	O75228	O75228 homo sapien
44	79	31.9	72	4	Q8TB48	Q8TB48 homo sapien
45	79	31.9	79	4	Q9UCY2	Q9UCY2 homo sapien

ALIGNMENTS

RESULT 1

Q7826 PRELIMINARY; PRT; 100 AA.
AC Q7826;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE X-linked retinopathy protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93224131; PubMed=8468040;
RA Wong P, MacDonald I M, Sood R, Smith C, Pilon R, Tenniswood M.;
RT Identification and partial characterization of a candidate gene for
RT X-linked retinopathies using a lateral approach.;
RL Genomics 15:467-471(1993).
DR EMBL; SS8722; AAB26149.1; -.
DR PIR; A46010; A46010.
FT NON_TER 1
SQ SEQUENCE 100 AA; 11065 MW; 50722631CC2CC479 CRC64;

Query Match 58.1%; Score 144; DB 4; Length 100;
Best Local Similarity 62.5%; Pred. No. 2.4e-11;
Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 1 FFFFLQSFLLSQAGVAWHDLGSLHPPLPGSSDSRASASQARITGV 48
Db 2 FFFFLQSFLLSQAGVAWHDLGSLHPPLPGSSDSRASASQARITGV 49

RESULT 2

Q8N181 PRELIMINARY; PRT; 109 AA.
AC Q8N181;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

```

DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DN OK/SW-CL.41.
GN OK/SW-CL.41.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Shchitjo S., Itoh K.;
RT "Identification of immuno-peptidomics that recognized by tumor-reactive
RT CTL generated from TIL of colon cancer patients.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062477; BAB93502.1; -.
SQ SEQUENCE 109 AA; 11580 MW; F55200153CEE2B33 CRC64;

Query Match 54.4%; Score 135; DB 4; Length 109;
Best Local Similarity 65.9%; Pred. No. 4.1e-10;
Matches 27; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 8 QSFILSQAGVAVHDLGSLHPLPGSSDSRASASOSARITGV 48
DB 2 ESHSVTQAGVQVHDLGSLHPLPGSSDSPTASERVAGITGM 42
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
Q9NX85 PRELIMINARY; PRT; 152 AA.
AC Q9NX85;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ20378.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Ileal mucosa;
RC Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000385; BAA91131.1; -.
KW Hypothetical protein.
SQ SEQUENCE 152 AA; 16568 MW; 59065F45AAA301B5 CRC64;

Query Match 52.8%; Score 131; DB 4; Length 152;
Best Local Similarity 38.0%; Pred. No. 2e-09;
Matches 35; Conservative 2; Mismatches 9; Indels 46; Gaps 2;

QY 2 FFFFLRQSFIL-SQAGVAVHDLGSLHPLPG- 31
DB 23 FFFFLRQSFALVAGVQVHDLGSLQPPPGFKQFCLSLSSWDYRHAPCPAYFVFLV 82
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 32 -----SSDRPASASOSARITG 47
DB 83 DMGPFHVQGTGLELLTSGDPPASASQAGITG 114
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
Q9NW14 PRELIMINARY; PRT; 208 AA.
AC Q9NW14;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ20837.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Adipose tissue;
RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Obayashi M., Nishi T.,
RA Shibahara T., Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000844; BAA91396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 208 AA; 23290 MW; C916648B9CAB4520 CRC64;

Query Match 50.0%; Score 124; DB 4; Length 208;
Best Local Similarity 65.7%; Pred. No. 2.5e-08;
Matches 23; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FFFFLRQSFILSQAGVAVHDLGSLHPLPGSSDS 35
DB 137 FFLFFDMESLSVAQAGVQVHDLGSLQPHLPGSSNS 171
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
Q9EGW3 PRELIMINARY; PRT; 122 AA.
AC Q9EGW3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]_TaxID=9541;
RP SEQUENCE FROM N.A.
RC TISSUE=Frontal cortex;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055293; BAB21918.1; -.
KW Hypothetical protein.
SQ SEQUENCE 122 AA; 13539 MW; 77EF2E874C60831 CRC64;

Query Match 49.4%; Score 122.5; DB 6; Length 122;
Best Local Similarity 35.2%; Pred. No. 2.1e-08;
Matches 32; Conservative 5; Mismatches 11; Indels 43; Gaps 2;

QY 1 FFFFLR-QSTLQAGVAVHDLGSLHPLPG- 31
DB 12 FFSFPLETESGVQAGVQVHDLGSLQPTPLGKQFCLSLPSSWNYRCLPPHLAKFCIF 71
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 32 -----SSDRPASASOSARITGV 48
DB 72 SRDYVQAGLELLTSGNPLASASQAGITGM 102
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
Q9N8C8 PRELIMINARY; PRT; 121 AA.
AC Q9N8C8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ39655.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```



```
SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK096974; BAC04918.1; -
KW Hypothetical protein.
SQ SEQUENCE 121 AA; 12917 MW; 4026598DE8735432 CRC64;

Query Match 49.2%; Score 122; DB 4; Length 121;
Best Local Similarity 55.5%; Pred. No. 2.4e-08;
Matches 26; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 3 FFFFLRQSFLLSQAGVAWHDLGSLHPPLPGSSDRASASQARITGV 48
DB 21 YLFGMEPHVAQAQVQWRDLGSLQPPPPGSRSPFASQVAGITGV 66

RESULT 7
Q8NAL9 PRELIMINARY; PRT; 162 AA.
ID Q8NAL9
AC Q8NAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ35525.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Spleen;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092844; BAC03988.1; -
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 17880 MW; E6D94CE17D0DC71 CRC64;

Query Match 48.8%; Score 121; DB 4; Length 162;
Best Local Similarity 55.3%; Pred. No. 4.6e-08;
Matches 26; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPPLPGSSDRASASQARITG 47
DB 89 FFFFEHPHVAQAQVQWRDLGSLQPPPPGSRSPFASQVAGITG 135

RESULT 8
Q8NAL9 PRELIMINARY; PRT; 158 AA.
ID Q8NAL9
AC Q8NAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furiya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yanazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092450; BAC03893.1; -
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 17414 MW; E151503015F2FE34 CRC64;

Query Match 47.6%; Score 118; DB 4; Length 158;
Best Local Similarity 33.0%; Pred. No. 1.1e-07;
Matches 31; Conservative 6; Mismatches 11; Indels 46; Gaps 2;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLH-----PPLPG----- 31
DB 54 FFFFEEKESLSTQAGVQWRDLGSLQAAAPPQFTPFSCLSLSPSSMNYRRPPCPANFFVFL 113

RESULT 9
Q9H387 PRELIMINARY; PRT; 118 AA.
ID Q9H387
AC Q9H387;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PRO2550.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Liver;
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1; -
SQ SEQUENCE 118 AA; 13257 MW; 94588870CAC8760D CRC64;

Query Match 47.2%; Score 117; DB 4; Length 118;
Best Local Similarity 36.2%; Pred. No. 1.1e-07;
Matches 34; Conservative 1; Mismatches 13; Indels 46; Gaps 2;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPPLP----- 30
DB 5 FFFFLRWSFTLVAQAGVQWRDLSSPPQPPPPKRFKFSCLSPSSWDRHAPPHPANFVFL 64

RESULT 10
Q8N874 PRELIMINARY; PRT; 138 AA.
ID Q8N874
AC Q8N874;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
```

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ39895.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hiraio M., Shimizu F.,
RA Wabe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097214; BAC04976.1; -
KW Hypothetical protein.
SQ SEQUENCE 138 AA; 15427 MW; 685645DF92E31994 CRC64;
Query Match 43.8%; Score 108.5; DB 4; Length 138;
Best Local Similarity 34.1%; Pred. No. 1.8e-06;
Matches 31; Conservative 4; Mismatches 11; Indels 45; Gaps 2;
QY 3 FFFLRQSTL-SQAGVAMHDLGSLHPLPG-----31
DB 13 FLFLSWSTLVAQTGVQWRNLGSLQPPPGPKRPSCLGLSSWDYRHHAWLIFVLLAEM 72
QY 32 -----SSDSRASASQARITGV 48
DB 73 GFCHVGQAGLELLTSSEPPASASQAGITGM 103

RESULT 11
Q8N845 PRELIMINARY; PRT; 157 AA.
AC Q8N845;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ40023.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stomach;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai K., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuna M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097342; BAC05007.1; -
KW Hypothetical protein.
SQ SEQUENCE 157 AA; 17930 MW; D79F77F3B7C608E CRC64;
Query Match 42.7%; Score 106; DB 4; Length 157;
Best Local Similarity 61.3%; Pred. No. 4.4e-06;
Matches 19; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 1 FFFFLRQSTL-SQAGVAMHDLGSLHPLPG 31
DB 86 FYFFFEESRSVAQAGVQWRDLGSLKPPSPG 116

RESULT 12
Q96EB1 PRELIMINARY; PRT; 535 AA.
AC Q96EB1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis, and Embryonic carcinoma;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012514; AAH2514.1; -
DR Genew; HGNC:1171; ELP4.
DR InterPro; IPR008728; PAXNEB.
DR Pfam; PF05625; PAXNEB; 1.
SQ SEQUENCE 535 AA; 58713 MW; 8656DD3B545B96D4 CRC64;
Query Match 42.5%; Score 105.5; DB 4; Length 535;
Best Local Similarity 57.8%; Pred. No. 2.1e-05;
Matches 26; Conservative 3; Mismatches 15; Indels 1; Gaps 1;
QY 3 FFFLRQSTL-SQAGVAMHDLGSLHPLPGSSDSRASASQARITG 47
DB 371 FLKLEKFTI-EAGVQWHDLSRRRLGSGGSPASASLVAGITG 414

RESULT 13
Q9H5D5 PRELIMINARY; PRT; 162 AA.
AC Q9H5D5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ23555.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027208; BAB15692.1; -
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 17632 MW; C9A857907E44D2B8 CRC64;
Query Match 42.3%; Score 105; DB 4; Length 162;
Best Local Similarity 61.3%; Pred. No. 6.1e-06;
Matches 19; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 1 FFFFLRQSTL-SQAGVAMHDLGSLHPLPG 31
DB 109 FFFFEEMSCVPAGVQWHDLSGQQPPPG 139

RESULT 14
Q9UI48 PRELIMINARY; PRT; 61 AA.
AC Q9UI48;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PRO0663 (Fragment)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 50 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF090944; AAF24056.1; -
FT NON TER 1
SQ SEQUENCE 61 AA; 7364 MW; AAFE987A88277368 CRC64;

Query Match 41.5%; Score 103; DB 4; Length 61;
Best Local Similarity 61.3%; Pred. No. 3.7e-06;
Matches 19; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPG 31
DB 3 FFFFLRQSFTLSQAGVAMHDLGSLHPPPG 33

RESULT 15
Q95KE1 PRELIMINARY; PRT; 135 AA.
AC Q95KE1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Frontal cortex;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062934; BAB60728.1; -
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.
DR GO; GO:0006036; P:glycolysis; IEA.
DR InterPro; IPR00172; GAP_dhhydrogenase.
DR PROSITE; PS00071; GAPDH; 1.
KW Hypothetical protein.
SQ SEQUENCE 135 AA; 15411 MW; 94DA9FF8025670DF CRC64;

Query Match 41.3%; Score 102.5; DB 6; Length 135;
Best Local Similarity 48.9%; Pred. No. 1.1e-05;
Matches 23; Conservative 3; Mismatches 10; Indels 11; Gaps 1;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPG 36
DB 77 FFFFLRQSFTLSQAGVAMHDLGSLHPPPG 123

Search completed: August 10, 2004, 19:15:15
Job time : 36 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:10:16 ; Search time 54 Seconds
(without alignments)
251.153 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSTLQAGVAWH.....LPSSDSRASASQARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	248	100.0	48	5	ABG65595	Abg65595 Human bre
2	160	64.5	110	4	AAU31552	AAU31552 Novel hum
3	160	64.5	113	4	AAU32788	AAU32788 Novel hum
4	159.5	64.3	105	4	AAU31652	AAU31652 Novel hum
5	153	61.7	176	7	ADC87115	ADC87115 Human GPC
6	153	61.7	217	7	ADC87151	ADC87151 Human GPC
7	152	61.3	114	4	AAU32787	AAU32787 Novel hum
8	147	59.3	85	4	AAU29593	AAU29593 Novel hum
9	144	58.1	70	3	AB28061	AB28061 Human sec
10	144	58.1	100	7	ADB37600	ADB37600 Neural th
11	143	57.7	112	4	AAU32786	AAU32786 Novel hum
12	140.5	56.7	194	4	AAU25539	AAU25539 Human pro
13	138.5	55.8	110	4	AAU30350	AAU30350 Novel hum
14	137.5	55.4	52	4	AAU11780	AAU11780 Human pol
15	137	55.2	151	7	ADC86929	ADC86929 Human GPC
16	136	54.8	145	4	AAU32025	AAU32025 Novel hum
17	135	54.4	97	5	ABP55149	ABP55149 Human lip
18	135	54.4	103	4	AAU33301	AAU33301 Novel hum
19	135	54.4	103	4	AAU33213	AAU33213 Novel hum
20	135	54.4	109	7	ADD22434	ADD22434 HLA-B46 T
21	135	54.4	164	4	AB117075	AB117075 Human ner
22	133.5	53.8	153	4	AAU31783	AAU31783 Novel hum
23	133	53.6	142	7	ADB64520	ADB64520 Human pro
24	132	53.2	156	4	ABG10943	ABG10943 Novel hum
25	132	53.2	293	4	AAU71968	AAU71968 Human bon

26	131.5	53.0	71	4	AAU33302	AAU33302 Novel hum
27	131.5	53.0	71	4	AAU33214	AAU33214 Novel hum
28	131.5	53.0	116	4	AAU14736	AAU14736 Novel bon
29	131	52.8	152	7	ADB37629	ADB37629 Neural th
30	131	52.8	164	4	AAU32072	AAU32072 Novel hum
31	131	52.8	229	4	AAU30354	AAU30354 Novel hum
32	130	52.4	53	4	AAO10159	AAO10159 Human pol
33	130	52.4	58	4	AAU95665	AAU95665 Human rep
34	130	52.4	67	4	AAO10596	AAO10596 Human pol
35	130	52.4	70	4	AAO2241	AAO2241 Human pol
36	130	52.4	101	4	AAU32073	AAU32073 Novel hum
37	130	52.4	103	4	AAO8294	AAO8294 Human pol
38	129.5	52.2	104	4	AAU33189	AAU33189 Novel hum
39	129.5	52.2	104	4	AAU33156	AAU33156 Novel hum
40	129.5	52.2	104	4	AAU33281	AAU33281 Novel hum
41	129	52.0	47	4	AAO10724	AAO10724 Human pol
42	129	52.0	117	7	ADD19326	ADD19326 Human sec
43	129	52.0	117	7	ADD19292	ADD19292 Human sec
44	129	52.0	138	4	AAU32881	AAU32881 Novel hum
45	128.5	51.8	324	4	AAU29573	AAU29573 Novel hum

ALIGNMENTS

RESULT 1
ABG65595
ID ABG65595 standard; protein; 48 AA.

XX AC ABG65595;
XX DT 23-AUG-2002 (first entry)
XX DE Human breast specific polypeptide, BSP #65.
XX KW Human; breast specific protein; BSP; breast cancer; vaccine;
XX KM mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.

XX OS Homo sapiens.
XX PN WO200236807-A2.
XX PD 10-MAY-2002.
XX PF 29-OCT-2001; 2001WO-US046888.
XX PR 27-OCT-2000; 2000US-0243802P.
XX (DIAD-) DIADEXUS INC.
XX Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
XX WPI; 2002-463415/49.

XX New breast-specific nucleic acids and polypeptides, useful for
XX identifying, diagnosing, monitoring, staging, imaging, and treating
XX breast cancer and non-cancerous disease states in breast tissues.
XX Claim 11; Page 245; 281pp; English.

XX The invention relates to breast specific nucleic acids (BSNA) and breast-specific polypeptides (BSP). Also included are a method for determining the BSNA in a sample, a vector comprising a BSNA, a host cell comprising the BSNA, a method for producing a polypeptide encoded by a BSNA, an anti-BSP antibody and a method for determining the presence of a BSP in a sample. The breast-specific nucleic acids, polypeptides and compositions comprising them are useful for identifying, diagnosing, monitoring, staging, imaging, and treating breast cancer, mammary tumour and non-cancerous disease states in breast tissue; for identifying breast tissue; for monitoring, identifying and/or designing agonists and antagonists of the polypeptides; in gene therapy; in producing transgenic animals and cells; for producing engineered breast tissue for treatment and research; and as elements in an array or computer program for pattern recognition

CC of breast disorders. The nucleic acids may be used as hybridisation
 CC probes to detect, characterise and quantify hybridising nucleic acids in,
 CC and isolate hybridising nucleic acids from, both genomic and transcript-
 CC derived nucleic acid samples. The BSP protein may be used in a vaccine
 CC composition for raising an immune response against breast cancer. The
 CC present sequence is BSP protein of the invention
 XX
 SQ Sequence 48 AA;

Query Match 100.0%; Score 248; DB 5; Length 48;
 Best Local Similarity 100.0%; Pred. No. 3.4e-25;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFFFLRQFTLSQAGVAMHDLGSLHPLPGSSDSRASQSAKITGV 48
 DB 1 FFFFLRQFTLSQAGVAMHDLGSLHPLPGSSDSRASQSAKITGV 48

RESULT 2
 AAU31552
 ID AAU31552 standard; protein; 110 AA.
 XX
 AC AAU31552;

DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #2043.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.
 XX WO200179449-A2.
 XX 25-OCT-2001.

PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX Claim 20; Page 457; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX Sequence 110 AA;

Query Match 64.5%; Score 160; DB 4; Length 110;
 Best Local Similarity 69.6%; Pred. No. 3.6e-13;
 Matches 32; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 2 FFFFLRQFTLSQAGVAMHDLGSLHPLPGSSDSRASQSAKITG 47
 DB 1 FFFFEVESRTVTQAGVQWHDJGSLQPPPPGSSDSPVSAHVAEITG 46

RESULT 3
 AAU32788
 ID AAU32788 standard; protein; 113 AA.

XX
 AC AAU32788;

DT 18-DEC-2001 (first entry)

XX Novel human secreted protein #3279.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.
 XX WO200179449-A2.
 XX 25-OCT-2001.

PF 16-APR-2001; 2001WO-US008656.

XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX Claim 20; Page 663-664; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX Sequence 113 AA;

Query Match 64.5%; Score 160; DB 4; Length 113;
 Best Local Similarity 69.6%; Pred. No. 3.7e-13;
 Matches 32; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 FFFFLRQFTLSQAGVAMHDLGSLHPLPGSSDSRASQSAKITG 47
 DB 31 FFFETESCSVAQAGVQWHDJGSLQPPPPGSSDSPASRVAGITG 76

```
RESULT 4
AAU31652
ID AAU31652 standard; protein; 105 AA.
XX
AC AAU31652;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2143.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
FN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008556.
XX
PR 18-APR-2000; 2000US-00552929.
XX
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
FI Tang YT, Liu C, Dmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 492; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention.
XX
SQ Sequence 105 AA;
Query Match 64.3%; Score 159.5; DB 4; Length 105;
Best Local Similarity 40.9%; Pred. No. 4e-13;
Matches 38; Conservative 2; Mismatches 8; Indels 45; Gaps 1;
QY 1 FFFFLRQSTLSQAGVAWHDLSLHPPLG----- 31
DB 1 FFFFLRQSFALQAQGWHDLSLQPPPGFKXFSYLSLPSWMDYRYVPPRPAPFEFLV 60
QY 32 -----SSDRASASQARITGV 48
DB 61 EMGFHYVQAGFELLTSSDPPASASQAGIIGV 93
RESULT 5
ADC87115
ID ADC87115 standard; protein; 176 AA.
XX
AC ADC87115;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1568.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
FN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
WPI; 2003-315783/31.
XX
DR N-PSDB; ADC87114.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 2; SEQ ID NO 1568; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ Sequence 176 AA;
Query Match 61.7%; Score 153; DB 7; Length 176;
Best Local Similarity 66.0%; Pred. No. 5.1e-12;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 FFFFLRQSTLSQAGVAWHDLSLHPPLGSSDRASASQARITG 47
DB 4 FFFFLRQSFHFTVQAGVQWCMNFSSLPPLGSSDSPASASQVAGITG 50
RESULT 6
ADC87151
ID ADC87151 standard; protein; 217 AA.
XX
AC ADC87151;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1604.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
FN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
```

PF 18-JUN-2002; 2002EP-00013517.
 XX
 PR 18-JUN-2001; 2001JP-00246789.
 XX
 XX (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ABSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 PI WPI; 2003-315783/31.
 XX N-PSDB; ADC87150.
 DR
 DR
 XX New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX
 XX Claim 2; SEQ ID NO 1604; 28pp; English.
 PS
 CC The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
 XX
 XX Sequence 217 AA;
 SQ
 Query Match 61.7%; Score 153; DB 7; Length 217;
 Best Local Similarity 68.1%; Pred. No. 6.5e-12;
 Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAVHDLGSLHPPLPGSSDSRASASQSRITG 47
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 146 PFFFEETESHVTWAGVQWCDLGLSLQPPPPGSSDSPASASRVAGITG 192
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 7
 ID AAU32787 standard; protein; 114 AA.
 XX
 AC AAU32787;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #3278.
 XX
 XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PT
 XX Claim 20; Page 663; 765pp; English.
 PS
 CC The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
 XX
 XX Sequence 217 AA;
 SQ
 Query Match 61.7%; Score 153; DB 7; Length 217;
 Best Local Similarity 68.1%; Pred. No. 6.5e-12;
 Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAVHDLGSLHPPLPGSSDSRASASQSRITG 47
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 146 PFFFEETESHVTWAGVQWCDLGLSLQPPPPGSSDSPASASRVAGITG 192
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 8
 ID AAU29593 standard; protein; 85 AA.
 XX
 AC AAU29593;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #84.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PT
 XX Claim 20; Page 168; 765pp; English.
 PS
 CC The invention relates to a novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 XX Sequence 114 AA;
 SQ
 Query Match 61.3%; Score 152; DB 4; Length 114;
 Best Local Similarity 66.7%; Pred. No. 4.3e-12;
 Matches 32; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAVHDLGSLHPPLPGSSDSRASASQSRITGV 48
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 38 PFFFEETESHCHVAGVQWDLGLSLQPPPPGSSNSPASASQVACTTGV 85
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 8
 ID AAU29593 standard; protein; 85 AA.
 XX
 AC AAU29593;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #84.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PT
 XX Claim 20; Page 168; 765pp; English.
 PS
 CC The invention relates to a novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 XX Sequence 114 AA;
 SQ
 Query Match 61.3%; Score 152; DB 4; Length 114;
 Best Local Similarity 66.7%; Pred. No. 4.3e-12;
 Matches 32; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAVHDLGSLHPPLPGSSDSRASASQSRITGV 48
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 38 PFFFEETESHCHVAGVQWDLGLSLQPPPPGSSNSPASASQVACTTGV 85
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 SQ Sequence 85 AA;

Query Match 59.3%; Score 147; DB 4; Length 85;
 Best Local Similarity 62.5%; Pred. No. 1.4e-11;
 Matches 30; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDRASASQSRITGV 48
 Db 5 YLFIVTESCTVVGQVQWCDLGLSLQPLPGSSDPHASTXVAGITGV 52

RESULT 9
 AAB28061
 ID AAB28061 standard; protein; 70 AA.
 XX
 AC AAB28061;

DT 02-FEB-2001 (first entry)
 DE Human secreted protein BLAST search protein SEQ ID NO: 109.

XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO2000055177-A2.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000WO-US006058.

XX 12-MAR-1999; 99US-0124145P.

XX 03-DEC-1999; 99US-0158654P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638177/61.

XX Novel nucleic acids encoding 49 human secreted proteins useful for
 PT treating cancers, hyperproliferative disorders, inflammatory disorders,
 PT neurological disorders and cardiovascular disorders.

XX Disclosure; Page 364-365; 389pp; English.

XX The invention relates to the isolation of genes AAC59108-C59156 encoding
 CC the human secreted proteins AAB28012-B28060. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene given in
 CC the descriptor line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital, (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX
 SQ Sequence 70 AA;

Query Match 58.1%; Score 144; DB 3; Length 70;
 Best Local Similarity 62.5%; Pred. No. 2.8e-11;
 Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDRASASQSRITGV 48
 Db 1 FFFPETESCVAEAGVQWCDLGLSKPPPGSSDSPAGASRVAGITGM 48

RESULT 10
 ADB37600
 ID ADB37600 standard; protein; 100 AA.
 XX
 AC ADB37600;

DT 04-DEC-2003 (first entry)

XX Neural thread protein-related protein #63.

XX Cytostatic; Antitumour; Antipsoriatic; Dermatological;

KW Antiatherosclerotic; Antiarteriosclerotic; Vasotropic; Antiinflammatory;

KW Immunosuppressive; Tranquillizer; Antiemetic; Virucide; AD7C-NTP;

KW neural thread protein; neuritic sprouting.

XX Unidentified.

XX WO2003008444-A2.

XX 30-JAN-2003.

XX 19-JUL-2002; 2002WO-CA001106.

XX 19-JUL-2001; 2001US-0306150P.

XX 19-JUL-2001; 2001US-0306161P.

XX 16-NOV-2001; 2001US-0331477P.

XX (NYMO-) NYMOX CORP.

XX Averbach PA, Gemmell J;

XX WPI; 2003-248000/24.

XX Novel Related peptide or AD7c-neural thread peptide, useful for treating
 PT unwanted cellular proliferations, glandular hyperplasia, unwanted facial
 PT hair, warts and unwanted fatty tissue.

XX Claim 1; Page 45; 109pp; English.

XX The present invention relates to AD7c-neural thread protein (NTP) and
 CC related proteins and peptides (I; ADB37528-ADB37641). The sequences are
 CC useful for treating a condition in a patient requiring removal or
 CC destruction of cells. The condition can be selected from benign or
 CC malignant tumour of a tissue, hyperplasia, hypertrophy or overgrowth of a
 CC tissue, vitally, bacterially or parasitically altered tissue, or
 CC malformation of a tissue, where the tissue is selected from lung, breast,
 CC stomach, pancreas, prostate, bladder, bone, ovary, skin, kidney, sinus,
 CC colon, intestine, stomach, rectum, oesophagus, heart, spleen, salivary
 CC gland, blood, brain and its coverings, spinal cord and its coverings,
 CC muscle, connective tissue, adrenal, parathyroid, thyroid, uterus, testis,
 CC pituitary, reproductive organs, liver, gall bladder, eye, ear, nose,
 CC throat, tonsils, mouth, lymph nodes and lymphoid tissue. The condition is
 CC preferably tonsillary hypertrophy, prostatic hyperplasia, psoriasis,
 CC eczema, dermatosis, cosmetic modification to a tissue (skin, eye, ear,
 CC nose, throat, mouth, muscle, connective, hair or breast tissue), vascular
 CC disease (atherosclerosis or arteriosclerosis), haemorrhoids, varicose

CC veins, inflammatory disease, autoimmune disease, metabolic disease,
 CC hereditary/genetic disease, traumatic disease or physical injury,
 CC nutritional deficiency disease, infectious disease, amyloid disease,
 CC fibrosis disease, storage disease, congenital malformation, enzyme
 CC deficiency disease, poisoning, intoxication, environmental disease,
 CC radiation disease, endocrine disease, degenerative disease and mechanical
 CC disease. The peptides are useful for treating unwanted cellular
 CC proliferations, glandular (e.g. prostate) hyperplasia, unwanted facial
 CC hair, warts and unwanted fatty tissue, or for preparing antibodies that
 CC recognize and/or bind to related proteins, Related peptides or NTP
 CC peptides.
 XX
 SQ Sequence 100 AA;

Query Match 58.1%; Score 144; DB 7; Length 100;
 Best Local Similarity 62.5%; Pred. No. 4.2e-11;
 Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
 QY 1 FFFFLRQFTLSQAGVAWHDLGSLHPLPGSSDSRASASQSRITGV 48
 DB 2 FFFFEETSCVAEAGVQWCDLGLKSPPGSSDSRASASRVAGITGM 49

RESULT 11
 AAU32786
 ID AAU32786 standard; protein; 112 AA.

XX AC AAU32786;
 XX DT 18-DEC-2001 (first entry)
 XX DE Novel human secreted protein #3277.
 XX KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX OS Homo sapiens.
 XX PN WO200179449-A2.
 XX PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US008656.

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX Claim 20; Page 663; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or

CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 SQ Sequence 112 AA;

Query Match 57.7%; Score 143; DB 4; Length 112;
 Best Local Similarity 65.9%; Pred. No. 6.4e-11;
 Matches 29; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 4 FFLRQFTLSQAGVAWHDLGSLHPLPGSSDSRASASQSRITG 47
 DB 1 FFEETSNVAQAGVQWHDGLGSLQPPSPGSSNPASASEVAGITG 44

RESULT 12
 AAU25539
 ID AAU25539 standard; protein; 194 AA.

XX AC AAU25539;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1054.

XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; antidiabetic; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antileuk; osteopathic; cytostatic;
 KW dermatological; antiallergic; antiaschmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX OS Homo sapiens.

XX PN WO200153455-A2.

XX PD 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US035017.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

XX N-PSDB; AAH99480.

XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX Claim 20; Page 212; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAU25225 to
 CC AAU25963. The proteins can have activities based on the tissues and cells
 CC they are expressed in, such as: antineoplastic; antirheumatic;
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
 CC antileuk; osteopathic; dermatological; antiallergic; antiaschmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;

CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders
 XX
 SQ Sequence 194 AA;

Query Match 56.7%; Score 140.5; DB 4; Length 194;
 Best Local Similarity 39.6%; Pred. No. 2.6e-10;
 Matches 36; Conservative 1; Mismatches 9; Indels 45; Gaps 2;

Qy 3 FFFLRQFTLL-SQAGVAVHDLGSLHPPPLPG----- 31
 Db 9 FFFLRWSFALVAQAGVQWHDLSLQPPAPGKRFSSLSLSLRWDYRHAHARLIFVFLVEM 68
 Qy 32 -----SSDSRASASQSARITGV 48
 Db 69 GFLHVGAGLELPTSGDPPTTSASQSARITGV 99

RESULT 13

AAU30350
 ID AAU30350 standard; protein; 110 AA.

XX
 AC AAU30350;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #841.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

XX
 EN WO200179449-A2.

XX
 PD 25-OCT-2001.

XX
 PF 16-APR-2001; 2001WO-US008656.

XX
 PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;

XX
 DR WPI; 2001-611725/70.

XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX
 PS Claim 20; Page 281; 765pp; English.

XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the

CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX
 SQ Sequence 110 AA;

Query Match 55.8%; Score 138.5; DB 4; Length 110;
 Best Local Similarity 58.5%; Pred. No. 2.5e-10;
 Matches 31; Conservative 5; Mismatches 12; Indels 5; Gaps 1;

Qy 1 FFFFL-----RQFTLSQAGVAVHDLGSLHPPPLPGSSDSRASASQSARITGV 48
 Db 13 FFFFLFFFTETESCFVQAQTQWCDLSLQPPPPGXNSPASASRVAGITGV 65

RESULT 14

AAO11780

ID AAO11780 standard; protein; 52 AA.

XX
 AC AAO11780;

XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human polypeptide SEQ ID NO 25672.

XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX
 FN WO200164835-A2.

XX
 PD 07-SEP-2001.

XX
 PF 26-FEB-2001; 2001WO-US004927.

XX
 PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577405.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;

XX
 DR WPI; 2001-514838/56.

DR N-PSDB; AAI91711.

XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.

XX
 PS Claim 20; SEQ ID NO 25672; 1399pp + Sequence Listing; English.

XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

Search completed: August 10, 2004, 19:14:04
Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:15:22 ; Search time 46 Seconds
(without alignments)
327.321 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSFLLSQAGVAMH.....LFGSSDRASQASARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 segs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09A_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	248	100.0	48	14	US-10-082-830-224
2	153	61.7	176	14	Sequence 224, App
3	153	61.7	176	15	Sequence 1912, App
4	153	61.7	217	15	Sequence 1568, App
5	153	61.7	217	14	Sequence 1956, App
6	144	58.1	217	15	Sequence 1604, App
7	140.5	56.7	194	12	Sequence 82, Appl
8	137	55.2	129	15	Sequence 1054, App
9	137	55.2	151	14	Sequence 4804, App
10	137	55.2	151	15	Sequence 1726, App
11	133	53.6	142	15	Sequence 1382, App
12	131	52.8	152	14	Sequence 2674, App
13	130	52.4	58	10	Sequence 112, App
14	126	50.8	220	14	Sequence 4323, App
15	126	50.8	220	15	Sequence 1698, App
					Sequence 1354, App

Sequence 1962, App
Sequence 115, App
Sequence 56, Appl
Sequence 68, Appl
Sequence 4053, App
Sequence 3256, App
Sequence 58, Appl
Sequence 674, App
Sequence 3047, App
Sequence 30, Appl
Sequence 1734, App
Sequence 1390, App
Sequence 1865, App
Sequence 908, App
Sequence 1959, App
Sequence 4725, App
Sequence 1850, App
Sequence 15, Appl
Sequence 5341, App
Sequence 107, App
Sequence 1940, App
Sequence 1980, App
Sequence 2008, App
Sequence 79, Appl
Sequence 3250, App
Sequence 3685, App
Sequence 2884, App
Sequence 2092, App
Sequence 1738, App

ALIGNMENTS

RESULT 1

US-10-082-830-224
; Sequence 224, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR FILING DATE: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 224
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-224

Query Match 100.0%; Score 248; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 6.1e-24;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAMHDLGSLHPLPGSSDRASQASARITGV 48
DB 1 FFFFLRQSFLLSQAGVAMHDLGSLHPLPGSSDRASQASARITGV 48

RESULT 2

US-10-017-161-1912
; Sequence 1912, Application US/10017161
; Publication No. US20030143668A1

```
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1912
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1912

Query Match          61.7%; Score 153; DB 14; Length 176;
Best Local Similarity 66.0%; Pred. No. 2.7e-11;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPLPGSSDSRASASQSRITG 47
      ||||| : : : : : : : : : : : : : : : : : :
DB 4 FFFFEESHVTVQAGVQWCMNFSSLPPLPGSSDSPASASQVAGITG 50

RESULT 3
US-10-292-798-1568
; Sequence 1568, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/156
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1568
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1568

Query Match          61.7%; Score 153; DB 15; Length 176;
Best Local Similarity 66.0%; Pred. No. 2.7e-11;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPLPGSSDSRASASQSRITG 47
      ||||| : : : : : : : : : : : : : : : : : :
DB 4 FFFFEESHVTVQAGVQWCMNFSSLPPLPGSSDSPASASQVAGITG 50

RESULT 4
US-10-017-161-1956
; Sequence 1956, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
```

```
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1956
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1956

Query Match          61.7%; Score 153; DB 14; Length 217;
Best Local Similarity 68.1%; Pred. No. 3.4e-11;
Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPLPGSSDSRASASQSRITG 47
      ||||| : : : : : : : : : : : : : : : : : :
DB 146 FFFFEESHVTVQAGVQWCDLGLSLQPPPGSSDSPASASRVAGITG 192

RESULT 5
US-10-292-798-1604
; Sequence 1604, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1604
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1604

Query Match          61.7%; Score 153; DB 15; Length 217;
Best Local Similarity 68.1%; Pred. No. 3.4e-11;
Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPLPGSSDSRASASQSRITG 47
      ||||| : : : : : : : : : : : : : : : : : :
DB 146 FFFFEESHVTVQAGVQWCDLGLSLQPPPGSSDSPASASRVAGITG 192

RESULT 6
US-10-198-070-82
; Sequence 82, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GEMMELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 59003.00008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
```


RESULT 13
US-09-764-
; Sequence
; Publicat
; GENERAL

APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1354
LENGTH: 220
TYPE: PRT
ORGANISM: Homo sapiens
US-10-292-798-1354

Query Match 50.8%; Score 126; DB 15; Length 220;
Best Local Similarity 56.5%; Pred. No. 9e-08;
Matches 26; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 2 FFFFLRQSFTLSQAGVAMHDLGSLHPPPLPGSSDSRASASQSRITG 47
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 44 YFFFLRQGLTVAQAGVQWCHGSLQPKTPGLSNPPTASQVAMTTG 89

Search completed: August 10, 2004, 19:20:51
Job time : 47 secs

This Page Blank (uspto)